1981

Analysis of ordered categorical responses, assuming an underlying continous variable

Robert Wayne Mee
Iowa State University

8128839

MEE, ROBERT WAYNE

ANALYSIS OF ORDERED CATEGORICAL RESPONSES, ASSUMING AN UNDERLYING CONTINUOUS VARIABLE

Iowa State University PH.D. 1981

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

Analysis of ordered categorical responses,

assuming an underlying continuous variable

by

Robert Wayne Mee

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1981

TABLE OF CONTENTS

GENERAL INTRODUCTION

Frequently, the response variable which is to assume the role of the dependent variable in a statistical analysis is not measured precisely. For example, calving difficulty in dairy cattle is sometimes recorded using the scale:

1 - No problem

2 - Slight problem

3 - Needed assistance

4 - Considerable force needed

5 - Extreme difficulty.

The term 'ordered categorical response' is used to refer to such a measurement. Note two characteristics of ordered categorical responses. First, although the categories are ordered, there is no 'unit of measurement' such as gallons of milk. Hence, the 'distance' between categories is not clearly defined. Second, the responses within a single category are not identical. For example, with calving difficulty, one birth may be more difficult than a second one, yet both may be recorded as 'No problem.'

Because of these two characteristics, linear models that may be appropriate for a continuous response variable are likely to be inappropriate for ordered categorical responses. [For exposition of the difficulties that arise from using an inappropriate model, see e.g., McKelvey and Zavoina (1975).] Suppose there are m response categories. One approach for modeling ordered categorical responses is to assume that there exist underlying continuous random variables and a partitioning of the real line into m intervals such that the i^{th} response is in Category j whenever the i^{th} underlying continuous variable falls in the j^{th} interval. This is referred to as the threshold approach. By assuming some model for the underlying continuous random variable, we may obtain a model for the observed responses which is consistent with the nature of ordered categorical responses.

Models based on the threshold approach have appeared frequently in the literature [e.g., Ashford (1959), Gurland, Lee, and Dahm (1960), Bock (1975), and McCullagh (1980)]. In each of these references, the underlying continuous random variable is assumed to satisfy a fixed effects linear model. However, the threshold approach is sufficiently versatile in that more general models for the underlying continuous variable could be assumed. McCullagh (1980) did consider the case of heteroscedastic error variances in the fixed effects model.

A primary objective of this thesis is to extend the threshold approach to the case where random as well as fixed effects are included in the underlying linear model. The use of this model is then illustrated by applying it to data on calving difficulty, where the objective is the prediction of the (random) sire effects and estimation of the ratio of the sire variance component to the error variance.

If, with calving difficulty, the integers 1, 2, ..., 5 (corresponding to the five categories mentioned previously) are used as the response variable in a mixed linear model analysis, several difficulties arise. Heteroscedastic variances in the 'observed scale' is one problem (Berger and Freeman, 1978). A second problem, noted e.g. by Tong, Wilton and

Schaeffer (1976), is that the estimated ratio of sire variance to error variance is determined in part by the proportion of difficult births observed. Berger and Freeman (1978) attempted to correct for heteroscedasticity by assuming a different error variance for each of 3 parity groups. However, this approach fails to accommodate either of the two characteristics of ordered categorical responses noted earlier. Tong, Wilton, and Schaeffer (1977) recommended the approach presented by Snell (1964) of assuming an underlying continuous variable, estimating the boundaries of the intervals on the underlying scale and, from these estimates, computing a score for each category. These scores are used as the response variable. While this approach allows for estimating the scores, it does not recognize that responses within a category are not necessarily identical. For this reason, Snell (1964, pg. 606) cautions "The use of scores for analyzing categorical data is justifiable only provided the grouping is not too severe .... Also, of course, there should be relatively few observations in the extreme scale categories." This caution surely applies to the analysis of calving difficulty, where more than 75% of the responses generally fall in the no difficulty class.

Geneticists have sometimes viewed categorical traits as the manifestation of an underlying continuous variable [see e.g. Falconer (1960, Chapter 18) and Gianola (1979)]. Therefore, the development of estimation procedures for a threshold model which assumes an underlying linear model with both fixed and random effects is of considerable interest.

Various other models for categorical responses have been proposed which include random effects. One approach is to express the response as a vector of indicator variables (with the k^{th} element equal to 1 if the response is in the k^{th} category, and zero otherwise) and to assume that these vectors follow a multivariate linear model (Landis and Koch, 1977). Quass and Van Vleck (1980) discussed the analysis of calving difficulty via this model. This approach assumes that the probability of falling in a particular category follows a linear model. Numerous constraints are required on the fixed and random effects to ensure that the estimated probabilities are contained in [0,1]. Alternatively, a log-linear model may be assumed for the probabilities. Laird (1975) assumes a log-linear model, with random effects, for the probabilities and uses the mode of the posterior distribution to estimate these random effects. Neither this model nor the one discussed by Quass and Van Vleck utilizes any ordering of the categories. This would seem to be a disadvantage when the categories have a meaningful order.

Thompson (1979) commented that investigation of a threshold-type model which includes random effects is needed. This dissertation presents such a model, together with a practical estimation procedure.

Explanation of Dissertation Format

The body of this dissertation is composed of three parts. In Parts I and II, the threshold approach is discussed assuming an underlying fixed effects linear model with independently and identically distributed random errors. These two parts present various topics including identifiability of the parameters, asymptotic properties of the maximum likelihood (ML) estimator, confidence intervals for probabilities, and iterative procedures for computing ML estimates. In Part III, the threshold approach assuming an underlying mixed effects linear model is presented and illustrated via the analysis of two data sets on calving difficulty. Equations are numbered by the part in which they appear, and then by the number within a part. For example, (3.1) indicates the first equation in Part III.

PART I.

ANALYSIS OF ORDERED CATEGORICAL RESPONSES, ASSUMING AN UNDERLYING CONTINUOUS VARIABLE

ABSTRACT

We consider inferences about the probability of an individual belonging to a particular one of m ordered categories. Whether an individual belongs to the j^{th} category is assumed to be determined by whether an underlying continuous random variable, which follows a linear model, falls within the j^{th} of m intervals whose endpoints are unknown. The resulting model for the probabilities is reparameterized so that the new parameters are identifiable. Conditions are given which ensure the asymptotic normality of the maximum likelihood estimator. Confidence intervals are derived for the probability of a response being in a particular category. The procedures are illustrated via an example.

KEY WORDS: Ordered categories; Threshold model; Fixed linear models; Multiple regression; Maximum likelihood estimation.

1. INTRODUCTION

Ordered categorical responses arise in many settings. For example, beef cattle are graded standard, good, choice, or prime. Each grade or category consists of individuals which are of similar but not identical "quality". Also, the grades have a natural ordering (with respect to quality).

Breed, age, and composition of diet could be considered as "predictor variables" for grade. Regression analysis is often used to make inferences about the relationship between a continuous response variable and various predictor variables. We consider the problem of making inferences about the relationship between an ordered categorical response and certain predictor variables. In particular, we consider inferences for the probability that the response will fall in a certain category or categories for given values of the predictor variables.

The approach taken is based on the assumption that the observed category is determined by the value of an underlying continuous random variable for which a linear model is applicable. Bock (1975, pg. 513) gave a clear description of this approach:

> "The underlying process is assumed scalar valued and distributed continuously in the population of subjects. There are assumed to be certain values on the continuum called thresholds, such that the m response categories correspond to intervals from $-\infty$ to $+\infty$ defined by the m-1 threshold values. The response of a given subject is determined by the interval in which his process value falls...."

Suppose that the underlying continuous random response variable y_i for the i^{th} of n "individuals" or "items" satisfies the linear model

$$y_i = \underline{x}_i\underline{\alpha} + e_i \qquad (i = 1,\ldots,n), \qquad (1.1)$$

where

$\underline{x}_i = (x_{i1},\ldots,x_{ip})$ is a known vector, whose elements represent those values of the independent or predictor variables associated with the i^{th} response,

$\underline{\alpha} = (\alpha_1,\ldots,\alpha_p)'$ is a vector of unknown parameters or fixed effects,

and e_i = a random error.

We assume that $e_1,\ldots,e_n$ are independently and identically distributed with mean 0 and strictly positive standard deviation σ. We denote by $\Phi(\cdot)$ the cumulative distribution function (CDF) of e_i/σ and assume that

$$\Phi(t) = \int_{-\infty}^{t} \phi(s)\, ds,$$

where $\phi(\cdot)$ is a known function that is continuous and strictly positive over the entire real line. Common choices for the error distribution are the standard normal distribution, where

$$\phi(t) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}t^2}, \qquad (1.2)$$

and the logistic distribution, where

$$\phi(t) = e^{-t}/(1-e^{-t})^2 , \qquad (1.3)$$

The underlying continuous variable y_i cannot be observed. We can only observe that the i^{th} response falls into a particular category. We number the m categories $1,2,\ldots,m$, and take Z_i to be the discrete random variable defined by $Z_i=j$ when the i^{th} individual or item belongs to Category j. The relationship bewteen y_i and Z_i is:

$$\xi_{j-1} < y_i \leq \xi_j \iff Z_i = j \ ,$$

where $j \in \{1,2,\ldots,m\}$, $\xi_0 = -\infty$, $\xi_m = \infty$, and $\xi_1,\ldots,\xi_{m-1}$ are unknown boundaries which define a partitioning of the real line into m intervals. Let $\underline{\theta} = (\xi_1,\ldots,\xi_{m-1}, \underline{\alpha}', \sigma)'$, and take the parameter space Θ to be

$$\Theta = \{\underline{\theta}:\xi_0 < \xi_1 \leq \cdots \leq \xi_{m-1} < \xi_m, \ \underline{\alpha} \in \underline{R}^p, \ \sigma > 0\} \ . \qquad (1.4)$$

Under these assumptions, the probability P_{ij} that $Z_i = j$, i.e., that the i^{th} individual or item belongs to Category j, is given by

$$\begin{aligned} P_{ij}(\underline{\theta}) &= \Pr(\xi_{j-1} < y_i \leq \xi_j) = \Phi[(\xi_j - \underline{x}_i\underline{\alpha})/\sigma] - \Phi[(\xi_{j-1} - \underline{x}_i\underline{\alpha})/\sigma] \\ &= \Phi[(1,\underline{x}_i)(1/\sigma)\binom{\xi_j}{-\underline{\alpha}}] - \Phi[(1,\underline{x}_i)(1/\sigma)\binom{\xi_{j-1}}{-\underline{\alpha}}] \ . \qquad (1.5) \end{aligned}$$

Note that $Z_1,\ldots,Z_n$ are independently distributed due to the assumed independence of the e_i's. Let $\underline{Z} = (Z_1,\ldots,Z_n)'$, and define $P(\cdot|\underline{\theta})$ to be the joint probability mass function of $\underline{Z}$, so that

$$P[(z_1,\ldots,z_n)'|\underline{\theta}] = \prod_{i=1}^{n} P_{iz_i}(\underline{\theta}) \quad (z_1=1,\ldots,m;\ldots,z_n=1,\ldots,m).$$

We refer to the above model for $Z_1,\ldots,Z_n$ as the threshold model. This model is appropriate for settings where:

1. The original responses are actually continuous, but (because of measurement limitations or for reasons of convenience) only categorical information is recorded; or underlying continuous responses can be conceptualized.

2. The linear model (1.1) would seem appropriate if we could actually observe the underlying continuous responses $y_1, \ldots, y_n$.

The threshold model is one of several models for ordered categorical responses. Aitchison and Silvey (1957) proposed another model which is defined in terms of underlying continuous random variables. Their model differs from the threshold model in that it includes m-1 independently distributed random variables for each individual (one for each threshold). Certain other methods for analyzing ordered categorical responses are based on the assumption that given functions of the probabilities $\Pr(Z_i=j)$ satisfy a linear model in a set of unknown parameters. Williams and Grizzle (1972) presented two such approaches. Their second approach, which is based on a model known as the scaling model, is, in certain special cases when $\Phi(\cdot)$ is taken to be a logistic CDF, identical to the threshold model.

The concept of relating an ordered categorical response to an underlying continuous variable appeared some time ago in the work of geneticists (e.g., Wright, 1934). Probit analysis (e.g., Finney, 1971) is based on a simple threshold model (that for the case $m=2$). Ashford (1959) was the first to discuss, from a linear model viewpoint, the threshold model when $m > 2$. Subsequently the threshold model has been discussed and applied by Gurland, Lee, and Dahm (1960), Snell (1964), McKelvey and Zavoina (1975), Bock (1975), and McCullagh (1980). Maximum likelihood (ML), with $\Phi(\cdot)$ taken to be either the normal or logistic CDF, has generally been used to estimate the parameters. McCullagh (1980) discussed various other possible choices for $\Phi(\cdot)$. He also considered the case where the standard deviation of e_i depends on $\underline{x}_i$. The desirability of the threshold approach relative to other methods has been discussed by McCullagh and by McKelvey and Zavoina (1975).

We consider the problem of inferences for the parametric function

$$\begin{aligned} P_{0j}(\underline{\theta}) &= \Phi[(\xi_j - \underline{x}_0\underline{\alpha})/\sigma] - \Phi[(\xi_{j-1} - \underline{x}_0\underline{\alpha})/\sigma] \qquad (1.6) \\ &= \Phi[(1,\underline{x}_0)(1/\sigma)\binom{\xi_j}{-\underline{\alpha}}] - \Phi[(1,\underline{x}_0)(1/\sigma)\binom{\xi_{j-1}}{-\underline{\alpha}}] , \end{aligned}$$

where $\underline{x}_0$ is a $1 \times p$ vector of constants. The quantity $P_{0j}(\underline{\theta})$ represents the probability that a future individual or item whose vector of independent variables equals $\underline{x}_0$ will fall in Category j.

Not all functions of the parameters of the threshold model are necessarily identifiable. In Section 2, we consider the question of identifiability and determine the conditions under which $P_{0j}(\underline{\theta})$ is an identifiable function [and thus the conditions under which meaningful inferences for $P_{0j}(\underline{\theta})$ are possible]. We further characterize the class of identifiable functions and then, in Section 3.1, reparameterize the model in terms of a second threshold model whose parameters are all identifiable. Previously McCullagh (1980) mentioned the problem of non-identifiability when discussing the uniqueness of ML estimates. He suggested that this problem is essentially the same as that of non-estimability in linear models and can be handled by adding constraints or via generalized inverses. Bock (1975) pointed out the need to reparameterize so that the model matrix, i.e., the $n \times p$ matrix whose rows are $\underline{x}_1, \ldots, \underline{x}_n$, has full column rank and does not include a column of 1's.

In Section 3.2, we give conditions under which the asymptotic distribution of the ML estimator is multivariate normal. In Section 4, we derive, on the basis of this asymptotic distribution, approximate confidence intervals for $P_{0j}(\underline{\theta})$. Previous work has focused on the point estimation of probabilities. In Section 5, we illustrate, in the context of an example studied by Ashford (1959), the confidence interval procedure and its potential usefulness.

2. IDENTIFIABILITY

The parameter vector $\underline{\theta}$ is not identifiable since distinct values of $\underline{\theta} \in \textcircled{H}$ do not necessarily give rise to distinct joint distributions for $Z_1, \ldots, Z_n$. For example, $\underline{\theta}=\underline{\theta}^*$ and $\underline{\theta}=2\underline{\theta}^*$ define the same joint distribution for $Z_1, \ldots, Z_n$. Hence, the data cannot help us distinguish between the values $\underline{\theta}^*$ and $2\underline{\theta}^*$.

While $\underline{\theta}$ itself is not identifiable, certain functions of $\underline{\theta}$ are identifiable. [A parametric function $g(\underline{\theta})$ is identifiable if, for all $\underline{\theta}^+$ and $\underline{\theta}^* \in \textcircled{H}$, $g(\underline{\theta}^+) \neq g(\underline{\theta}^*)$ implies that $P(\cdot|\underline{\theta}^+)$ is not the same as $P(\cdot|\underline{\theta}^*)$.]

<u>Lemma 1</u>: Under the threshold model, $(\xi_j - \underline{x}_i\underline{\alpha})/\sigma$ is identifiable $(i = 1,\ldots,n,\ j = 1,\ldots,m-1)$.

<u>Proof</u>: It suffices to show that for all $\underline{\theta}^+$ and $\underline{\theta}^* \in \textcircled{H}$, $P(\underline{z}|\underline{\theta}^+)$

$= P(\underline{z}|\underline{\theta}^*)$ for all $\underline{z}$ implies $(\xi_j^+ - \underline{x}_i\underline{\alpha}^+)/\sigma^+ = (\xi_j^* - \underline{x}_i\underline{\alpha}^*)/\sigma^*$, where $\underline{\theta}^+ = (\xi_1^+,\ldots,\xi_{m-1}^+, \underline{\alpha}^{+\prime},\sigma^+)'$ and $\underline{\theta}^* = (\xi_1^*,\ldots,\xi_{m-1}^*, \underline{\alpha}^{*\prime},\sigma^*)'$. Suppose

$$P(\underline{z}|\underline{\theta}^+) = P(\underline{z}|\underline{\theta}^*) \text{ for all } \underline{z} . \tag{1.7}$$

Then, $\Pr(Z_i \leq j)$ is the same when $\underline{\theta} = \underline{\theta}^+$ as when $\underline{\theta} = \underline{\theta}^*$, which implies that

$$\Phi[(\xi_j^+ - \underline{x}_i\underline{\alpha}^+)/\sigma^+] = \Phi[(\xi_j^* - \underline{x}_i\underline{\alpha}^*)/\sigma^*] ,$$

and thus, since Φ is strictly increasing that

$$(\xi_j^+ - \underline{x}_i\underline{\alpha}^+)/\sigma^+ = (\xi_j^* - \underline{x}_i\underline{\alpha}^*)/\sigma^* .$$

This completes the proof.

Since a function of identifiable functions is itself identifiable and since the probability distribution of $\underline{Z}$ depends on the value of $\underline{\theta}$ only through the value of the functions $(\xi_j - \underline{x}_i\underline{\alpha})/\sigma$ $(i = 1,\ldots,n;$ $j = 1,\ldots,m-1)$, we have the following corollary to Lemma 1.

Corollary 1: Under the threshold model, a parametric function is identifiable if and only if it can be expressed as a function of the parametric functions $(\xi_j - \underline{x}_i\underline{\alpha})/\sigma$ $(i = 1,\ldots,n;\ j = 1,\ldots,m-1)$.

Define $\underline{X}$ to be the $n \times p$ matrix whose i^{th} row is $\underline{x}_i$, take $\underline{1}$ to be a column vector of 1's, and denote the row space of any matrix $\underline{A}$ by $R(\underline{A})$.

Corollary 2: Under the threshold model, the probability $P_{0j}(\underline{\theta})$ defined by (1.6) is identifiable if and only if $(1,\underline{x}_0) \in R(\underline{1},\underline{X})$.

Proof: Suppose $(1,\underline{x}_0) \varepsilon R(\underline{1},\underline{X})$. Then, there exists a vector $\underline{a}$ such that $\underline{a}'(\underline{1},\underline{X}) = (1,\underline{x}_0)$. It follows from Corollary 1 that

$$P_{0j}(\underline{\theta}) = \Phi[\underline{a}'(\underline{1}\xi_j - \underline{X\alpha})/\sigma] - \Phi[\underline{a}'(\underline{1}\xi_{j-1}-\underline{X\alpha})/\sigma]$$

is identifiable.

Conversely, suppose $(1,\underline{x}_0) \notin R(\underline{1},\underline{X})$. Take $\underline{\theta}^+ = (\xi_1^+,\ldots,\xi_{m-1}^+, \alpha_1^+,\ldots,\alpha_p^+, \sigma^+)$ to be a value of $\underline{\theta}$ such that $\xi_{j-1}^+ < \xi_j^+$. There exists a constant c such that $\Phi(\lambda_j) - \Phi(\lambda_{j-1}) \neq \Phi(\lambda_j + c) - \Phi(\lambda_{j-1} + c)$, where $\lambda_i = (\xi_i^+ - \underline{x}_0\underline{\alpha}^+)/\sigma^+$. Also, there exists a vector $\underline{b} = (b_0,\ldots,b_p)'$ such that $(1,\underline{x}_0)\underline{b} = c\sigma^+$ but $(\underline{1},\underline{X})\underline{b} = \underline{0}$, where $\underline{0}$ is a vector of 0's. Taking $\underline{\theta}^* = (\xi_1^+ + b_0,\ldots,\xi_{m-1}^+ + b_0, \alpha_1^+ - b_1,\ldots,\alpha_p^+ - b_p,\sigma^+)'$, we have that $P_{0j}(\underline{\theta}^+) \neq P_{0j}(\underline{\theta}^*)$ but $P(\underline{z}|\underline{\theta}^+) = P(\underline{z}|\underline{\theta}^*)$ for all $\underline{z}$. Hence, $P_{0j}(\underline{\theta})$ is not identifiable, and the proof is complete.

The probability mass function $P(\cdot|\underline{\theta})$ can be re-expressed in terms of "non-redundant" identifiable functions of $\underline{\theta}$. Define $r = \text{rank}(\underline{1},\underline{X})-1$, and let $\underline{W}$ represent any $n \times r$ matrix such that

$$M(\underline{1},\underline{W}) = M(\underline{1},\underline{X}), \tag{1.8}$$

where $M(\underline{A})$ denotes the column space of any matrix $\underline{A}$. The martix $(\underline{1},\underline{W})$ necessarily has full column rank.

Since the function $(\xi_j - \xi_1)/\sigma$ equals $(\xi_j - \underline{x}_i\underline{\alpha})/\sigma - (\xi_1 - \underline{x}_i\underline{\alpha})/\sigma$, it is identifiable (Corollary 1). Define parametric functions $\eta_1(\underline{\theta}),\ldots,\eta_{m-1}(\underline{\theta})$ and $\underline{\tau}(\underline{\theta}) = [\tau_1(\underline{\theta}),\ldots,\tau_r(\underline{\theta})]'$ by

$$\underline{1}\eta_1(\underline{\theta}) - \underline{W}\,\underline{\tau}(\underline{\theta}) = \sigma^{-1}(\underline{1}\xi_1 - \underline{X\alpha}) \tag{1.9}$$

and

$$\eta_j(\underline{\theta}) - \eta_1(\underline{\theta}) = (\xi_j - \xi_1)/\sigma \quad (j = 2,\ldots, m-1). \tag{1.10}$$

Note that condition (1.8) implies the existence of a unique solution to (1.9), so that the functions $\eta_1(\underline{\theta}),\ldots,\eta_{m-1}(\underline{\theta})$, $\tau_1(\underline{\theta}),\ldots,\tau_r(\underline{\theta})$ are well-defined. Further, it follows from Corollary 1 that $\eta_1(\underline{\theta}),\ldots,$ $\eta_{m-1}(\underline{\theta})$, $\tau_1(\underline{\theta}),\ldots,\tau_r(\underline{\theta})$ are identifiable functions.

Since

$$\underline{1}\eta_j(\underline{\theta}) - \underline{W}\ \underline{\tau}(\underline{\theta}) = \sigma^{-1}(1\xi_j - \underline{X\alpha}), \tag{1.11}$$

(1.5) may be expressed as

$$P_{ij}(\underline{\theta}) = \Phi[\eta_j(\underline{\theta}) - \underline{w}_i\underline{\tau}(\underline{\theta})] - \Phi[\eta_{j-1}(\underline{\theta}) - \underline{w}_i\underline{\tau}(\underline{\theta})], \tag{1.12}$$

where $\underline{w}_i$ denotes the i^{th} row of $\underline{W}$ and $\eta_0(\underline{\theta})$ and $\eta_m(\theta)$ denote $-\infty$ and $+\infty$, respectively. Similarly, if $P_{0j}(\theta)$ is identifiable, or equivalently if there exists a vector $\underline{a}$ such that $(\underline{1},\underline{x}_0) = \underline{a}'(\underline{1},\underline{X})$, then (1.6) may be reexpressed as

$$P_{0j}(\underline{\theta}) = \Phi[\eta_j(\underline{\theta}) - \underline{w}_0\underline{\tau}(\underline{\theta})] - \Phi[\eta_{j-1}(\underline{\theta}) - \underline{w}_0\underline{\tau}(\underline{\theta})] , \tag{1.13}$$

where

$$\underline{w}_0 = \underline{a}'\underline{W} . \tag{1.14}$$

We conclude this section by displaying the matrix $\underline{W}$ for two simple examples.

<u>Example 1</u>: Suppose that the model for the underlying continuous responses consists of a simple linear regression without an intercept, i.e., that $y_i = \alpha_1 x_i + e_i$. Then $\underline{X} = (x_1,\ldots,x_n)'$ and $(\underline{1},\underline{X})$ has full

column rank (provided of course that $x_i \neq x_{i'}$ for some i and i'). We may choose $\underline{W} = \underline{X}$, in which case $[\eta_1(\underline{\theta}),\ldots,\eta_{m-1}(\underline{\theta}),\tau_1(\underline{\theta})] = (\xi_1/\sigma,\ldots,\xi_{m-1}/\sigma,\ \alpha_1/\sigma)$.

Example 2: Suppose that the model for $y_1,\ldots,y_n$ is a simple linear regression with an intercept, i.e., that $y_i = \alpha_1 + \alpha_2 x_i + e_i$. Then r=1 and we may choose $\underline{W} = (x_1,\ldots,x_n)'$, in which case $[\eta_1(\underline{\theta}),\ldots,\eta_{m-1}(\underline{\theta}),\tau_1(\underline{\theta})] = [(\xi_1-\alpha_1)/\sigma,\ldots,(\xi_{m-1}-\alpha_1)/\sigma,\alpha_2/\sigma]$.

3. MAXIMUM LIKELIHOOD ESTIMATION

The likelihood function (ℓ) for the threshold model can be expressed as

$$\ell(\underline{\theta}) = \prod_{j\varepsilon C}\ \prod_{i\varepsilon R_j} P_{ij}(\underline{\theta}) \ ,$$

where R_j = {i: observed category of the i^{th} item or individual is Category j; i=1,...,n}(j = 1,...,m) and C = {j: $R_j \neq$ empty set}. The ML estimate of $\underline{\theta}$ (if it exists) is defined to be an element $\hat{\underline{\theta}}$ of Ⓗ such that

$$\ell(\hat{\underline{\theta}}) = \sup_{\underline{\theta}\varepsilon Ⓗ} \ell(\underline{\theta}) \ . \tag{1.15}$$

The log-likelihood function (L) is given by

$$L(\underline{\theta}) = \sum_{j\varepsilon C}\ \sum_{i\varepsilon R_j} \ln[P_{ij}(\underline{\theta})].$$

If $\hat{\underline{\theta}}$ is in the interior of Ⓗ, i.e., in Ⓗ^{o} = $\{\underline{\theta}:\xi_0 < \xi_1 < \ldots < \xi_{m-1} < \xi_m,\ \underline{\alpha}\varepsilon \underline{R}^p,\ \sigma > 0\}$, then it satisfies the likelihood equations

$$\partial L/\partial\theta_u\Big|_{\underline{\theta}=\hat{\underline{\theta}}} = 0 \quad (u = 1,\ldots,m+p) \ ,$$

where $\underline{\theta} = (\theta_1,\ldots,\theta_{m+p})' = (\xi_1,\ldots,\xi_{m-1},\ \alpha_1,\ldots,\alpha_p,\ \sigma)'$.

Haberman (1980) gave conditions for the existence of ML estimates for the threshold model. He stated that a necessary condition (when the parameter space is restricted to Θ^o) is that each category $(j = 1,\ldots,m)$ contain at least one response.

Suppose that $R_{j'}$ is empty, i.e., that there are no items or individuals in Category j', where $2 \leq j' \leq m-1$, and suppose that R_j is nonempty for some $j > j'$. Let Category j'' denote the first category after Category j' that contains at least one individual. Take $\underline{\theta}^+$ to be a value of $\underline{\theta}$ for which $\xi^+_{j'-1} < \xi^+_{j'}$, and take the elements of $\underline{\theta}^* = (\xi^*_1,\ldots,\xi^*_{j-1}, \underline{\tau}^{*\prime},\sigma^*)'$ to be equal to those of $\underline{\theta}^+$ except take $\xi^*_{j'} = \ldots = \xi^*_{j''-1} = \xi^+_{j'-1}$. Thus, $\underline{\theta}^*$ corresponds to $\underline{\theta}^+$ with Categories j' through $(j''-1)$ collapsed. Then,

$$
\begin{aligned}
L(\underline{\theta}^*) - L(\underline{\theta}^+) &= \sum_{j\in C}\ \sum_{i\in R_j} [\ln\{P_{ij}(\underline{\theta}^*)\} - \ln\{P_{ij}(\underline{\theta}^+)\}] \\
&= \sum_{i\in R_{j''}} [\ln\{P_{ij''}(\underline{\theta}^*)\} - \ln\{P_{ij''}(\underline{\theta}^+)\}] \\
&> 0 ,
\end{aligned}
$$

since $P_{ij''}(\underline{\theta}^*) = P_{ij'}(\underline{\theta}^+) + \ldots + P_{ij''}(\underline{\theta}^+) > P_{ij''}(\underline{\theta}^+)$ for all i. We conclude that, if L attains a maximum for $\underline{\theta}\in\Theta$, it must be at a point for which $\xi_{j'-1} = \xi_{j'}$.

If R_1 or R_m is empty, i.e., if there are no items or individuals in the first category or the last category, then an ML estimate does not exist. Let j^o = the largest element in C, and suppose $j^o < m$. Then, $L(\underline{\theta})$ is a strictly increasing function of ξ_{j^o} and so, $\xi_{j^o} = \ldots = \xi_{m-1}$

may be taken arbitrarily large. Similarly, let j_o = the smallest element in C. If $j_o > 1$, then $L(\underline{\theta})$ is a strictly decreasing function of ξ_{j_o-1} and $\xi_1 = \ldots = \xi_{j_o-1}$ may be taken arbitrarily small.

We conclude that ML will in effect collapse any category which contains no observations, and for purposes of computing an ML estimate of $\underline{\theta}$, we can adopt a strategy of deleting the categories that contain no observations, decreasing m accordingly and renumbering the remaining categories.

Anderson (1981) reported that, when he applied ML for one example with 4 categories, 5 treatments, and 20 blocks, six of the block estimates were unbounded. Whenever the responses z_i in a block are all in the first (or last) category, that block estimate will not exist, since the likelihood function is a monotone function of that estimate. If that block effect is taken sufficiently large negative (positive), $P_{iz_i}(\underline{\theta}) \doteq 1$ for all i in that block. Thus, for purposes of computing an ML estimate, we may delete all observations from blocks for which all responses are in Category 1 or all are in Category m.

Take $\underline{\omega}(\underline{\theta})$ to be the transformation $\underline{\omega}(\underline{\theta}) = [\eta_1(\underline{\theta}),\ldots,\eta_{m-1}(\underline{\theta}), \underline{\tau}(\underline{\theta})']'$, whose components are defined by (1.9) and (1.10). When an ML estimate $\hat{\underline{\theta}}$ exists, it will not be unique (since the elements of $\underline{\theta}$ are not identifiable). However, for $\hat{\underline{\theta}}$ satisfying (1.15), $\underline{\omega}(\hat{\underline{\theta}})$ is unique provided that $\ln[\phi(\cdot)]$ is concave (see Haberman, 1980). [For $\phi(\cdot)$ given by (1.2) or (1.3), $\ln[\phi(\cdot)]$ is concave.] We now introduce an alternative threshold model with identifiable parameters and show that inferences for identifiable functions of $\underline{\theta}$, when based on ML, are equivalent for the original and alternative models.

3.1 Alternative Threshold Model

Consider an alternative threshold model in which the underlying continuous variables, say $t_1,\ldots,t_n$, satisfy the linear model

$$t_i = \underline{w}_i\underline{\tau} + d_i \qquad (i = 1,\ldots,n), \tag{1.16}$$

where $\underline{\tau} = (\tau_1,\ldots,\tau_r)'$ is a vector of unknown parameters, and $d_1,\ldots,d_n$ are independently distributed random errors with common CDF Φ (and $\underline{w}_i$ is as defined in Section 2). Suppose that the relationships between t_i and the categorical response Z_i is taken to be

$$\eta_{j-1} < t_i \leq \eta_j \iff Z_i = j \ ,$$

where $j\varepsilon\{1,\ldots,m\}$, $\eta_0 = -\infty$, $\eta_m = +\infty$, and $\eta_1,\ldots,\eta_{m-1}$ are the unknown boundaries for the alternative threshold model. Define $\underline{\omega}$ as $\underline{\omega} = (\omega_1,\ldots,\omega_s)' = (\eta_1,\ldots,\eta_{m-1},\underline{\tau}')'$, where $s = m-1+r$. The parameter space Ω for $\underline{\omega}$ is taken to be

$$\Omega = \{\underline{\omega}:\eta_0 < \eta_1 \leq \cdots \leq \eta_{m-1} < \eta_m, \ \underline{\tau}\varepsilon\underline{R}^r\}. \tag{1.17}$$

Under the alternative threshold model, the probability that $Z_i = j$ is

$$P^*_{ij}(\underline{\omega}) = \Pr(\eta_{j-1} < t_i \leq \eta_j) = \Phi(\eta_j - \underline{w}_i\underline{\tau}) - \Phi(\eta_{j-1} - \underline{w}_i\underline{\tau}). \tag{1.18}$$

We define $P^*(\cdot|\underline{\omega})$ to be the joint probability mass function for $\underline{Z}$ under the alternative threshold model, i.e.,

$$P^*[(z_1,\ldots,z_n)'\,|\,\underline{\omega}] = \prod_{i=1}^{n} P^*_{iz_i}(\underline{\omega}) \quad (z_1=1,\ldots,m;\ldots;z_n=1,\ldots,m).$$

From (1.12) and (1.18) we have that $P^*[\underline{z}|\underline{\omega}(\underline{\theta})] = P(\underline{z}|\underline{\theta})$ for all $\underline{z}$. Furthermore, the parameter spaces Θ and Ω are equivalent in the sense that the set of possible values of $\underline{\omega}(\underline{\theta})$ under the original model, i.e., $\{\underline{\omega}(\underline{\theta}):\underline{\theta}\epsilon\Theta\}$, coincides with the parameter space Ω under the alternative model, implying in particular that $\underline{\omega}$ is identifiable.

Let $\hat{\underline{\omega}} = (\hat{\eta}_1,\ldots,\hat{\eta}_{m-1}, \hat{\underline{\tau}}')'$ represent an ML estimate of the parameter vector $\underline{\omega}$ of the alternative threshold model. Due to the invariance property of ML estimates, $\hat{\underline{\omega}}$ is an ML estimate of the vector $\underline{\omega}(\underline{\theta})$ of parametric functions of the parameter vector $\underline{\theta}$ of the original threshold model. Further, any identifiable function $g(\underline{\theta})$ can be expressed as $h[\underline{\omega}(\underline{\theta})]$ for some h [refer to Corollary 1 and (1.11)], and $h(\hat{\underline{\omega}})$ is an ML estimate of $g(\underline{\theta})$. Thus, in computing an ML estimate of $g(\underline{\theta})$, we can work with the alternative threshold model.

Consider in particular the function $P_{0j}(\underline{\theta})$ defined in (1.6). If $P_{0j}(\underline{\theta})$ is identifiable, then, defining

$$P^*_{0j}(\underline{\omega}) = \Phi(\eta_j - \underline{w}_0\underline{\tau}) - \Phi(\eta_{j-1} - \underline{w}_0\underline{\tau}) \tag{1.19}$$

[with $\underline{w}_0$ as in (1.14)], we have, from (1.13) and (1.19), that $P^*_{0j}[\underline{\omega}(\underline{\theta})] = P_{0j}(\underline{\theta})$. Thus, $P^*_{0j}(\hat{\underline{\omega}})$ is an ML estimate of $P_{0j}(\underline{\theta})$.

3.2 Asymptotic Properties of the ML Estimator

In this section, we establish sufficient conditions for the ML estimator of the parameter vector $\underline{\omega}$ of the alternative threshold model to be consistent and asymptotically normal. We do so by applying

Hoadley's (1971) Theorems 1 and 2.

Let $t_1, t_2, t_3, \ldots$ represent an infinite sequence of underlying continuous response variables such that, for any n ($n \geq k$ for some integer k),

$$t_1, t_2, \ldots, t_n \tag{1.20}$$

satisfy the alternative threshold model (1.16). Let $Z_1, Z_2, \ldots$ and $\underline{w}_1, \underline{w}_2, \ldots$ represent the corresponding sequences of categorical responses and vectors of independent variables. For each n, let $(\underline{1}_n, \underline{W}_n)$ denote the $n \times (r+1)$ matrix with i^{th} row $(1, \underline{w}_i) = (1, w_{i1}, \ldots, w_{ir})$, let $L_n(\cdot)$ denote the log-likelihood function, and let $\hat{\underline{\omega}}_n$ represent the ML estimator of $\underline{\omega}$. Finally, for $i = 1,2,\ldots$, define an $s \times s$ matrix $B_i(\underline{\omega})$ by

$$B_i(\underline{\omega}) = -E\{\partial^2 \ln[P^*_{iZ_i}(\underline{\omega})]/\partial\underline{\omega}\partial\underline{\omega}'\}.$$

__Theorem 1__: Consider, for $n = k, k+1, \ldots$, the alternative threshold model with underlying response variables given by (1.20). Let $\underline{\omega}^o = (\eta_1^o, \ldots, \eta_{m-1}^o, \underline{\tau}^o)' \varepsilon \Omega$ represent the true value of $\underline{\omega}$. Suppose that $\ln[\phi(\cdot)]$ is a concave function. Suppose further that:

1. There exists a constant M such that $|w_{iu}| \leq M$, ($i = 1,2,\ldots$; $u = 1,\ldots, r$).
2. There exists a positive definite matrix $\underline{D}$ such that

$$\underline{D} = \lim_{n\to\infty} n^{-1}(\underline{1}_n, \underline{W}_n)'(\underline{1}_n, \underline{W}_n). \tag{1.21}$$

Then, $\hat{\underline{\omega}}_n$ converges in probability ($\to$p) to $\underline{\omega}^o$.

<u>Proof</u>: Take K to be a constant such that the absolute value of each element of $\underline{\omega}^o$ is less than K. Define

$$\Omega^K = \{\underline{\omega}: -K \leq \eta_1 \leq \dots \leq \eta_{m-1} \leq K \; ; \; |\tau_u| \leq K(u=1,\dots,r)\}.$$

For each n, let $\overline{\underline{\omega}}_n \varepsilon \Omega^K$ denote that ML estimator of $\underline{\omega}$ when the parameter space is taken to be Ω^K. First we show that $\overline{\underline{\omega}}_n \to p \; \underline{\omega}^o$ by verifying that the conditions [Hoadley's conditions C1, C2, C3(i), C3(ii), C4(i), C4(ii), and C5] of Hoadley's Theorem 1 are satisfied. Then, we show that $\hat{\underline{\omega}}_n - \overline{\omega}_n \to p \; \underline{0}$, and thus that $\hat{\underline{\omega}}_n \to p \; \underline{\omega}^o$.

Hoadley's conditions C1, C2, and C5 are easy to verify and his conditions C3(ii) and C4(ii) are trivially satisfied since the parameter space Ω^K is compact. It remains to show that his conditions C3(i) and C4(i) are satisfied.

Define the random variables $R_i(\underline{\omega})(i=1,2,\dots)$ as

$$R_i(\underline{\omega}) = \left\{\begin{array}{ll} \ln[P^*_{iZ_i}(\underline{\omega})/P^*_{iZ_i}(\underline{\omega}^o)] & \text{if } P^*_{iZ_i}(\underline{\omega}^o) > 0 \\ 0 & \text{otherwise} \end{array}\right\}.$$

To show that Hoadley's condition C3(i) is satisfied, it suffices to show that $R_i(\underline{\omega})$ is bounded from above. Let $M^o = \text{Maximum}[\,|\eta_j^o - \underline{w}_i\underline{\tau}^o|:$ $i=1,\dots; j=1,\dots,m-1]$, let $\delta = \text{Minimum}\;[\eta_j^o - \eta_{j-1}^o\text{: j such that } \eta_j \neq \eta_{j-1}^o]$ and let $\gamma = \text{Infinum}[\phi(t): |t| \leq M^o + \delta]$. Since $\phi(t)$ is continuous, $\phi(c) = \gamma$ for some $c \;\varepsilon\; [-M^o-\delta, M^o+\delta]$ and thus $\gamma > 0$. Therefore, $\ln[1/(\delta\gamma)]$ is an upper bound for $R_i(\underline{\omega})$.

Let $\underline{\omega}^+ = (\eta_1^+,\ldots,\eta_{m-1}^+, \underline{\tau}^{+}{}')'$ be an arbitrary vector in $\Omega^K - \{\underline{\omega}^o\}$. To show that Hoadley's condition C4(i) is satisfied, it suffices to show that

$$\lim_{n\to\infty}\sup \frac{1}{n}\sum_{i=1}^{n} E[R_i(\underline{\omega}^+)] < 0 \tag{1.22}$$

(where the expectation is taken at $\underline{\omega} = \underline{\omega}^o$). We have that

$$E[R_i(\underline{\omega}^+)] = \sum_{j\epsilon C^o} g[P^*_{ij}(\underline{\omega}^+),\ P^*_{ij}(\underline{\omega}^o)], \tag{1.23}$$

where $g(a,b) = b[\ln(a/b)]$ and $C^o = \{j: P^*_{ij}(\underline{\omega}^o) > 0\}$. It can be readily verified that $g(\cdot,\cdot)$ is concave and thus, that

$$g(a_1,b_1) + g(a_2,b_2) \leq 2g[(a_1+a_2)/2, (b_1+b_2)/2] = g(a_1+a_2,\ b_1+b_2). \tag{1.24}$$

Since $P^*_{i1}(\underline{\omega}) + \ldots + P^*_{ij'}(\underline{\omega}) = \Phi(\eta_{j'}, -\underline{w}_i\underline{\tau})$ and $P^*_{ij'+1}(\underline{\omega}) + \ldots + P^*_{im}(\underline{\omega}) = 1 - \Phi(\eta_{j'}, -\underline{w}_i\underline{\tau})$, using (1.24) we find by using (1.24) that

$$E[R_i(\underline{\omega}^+)] \leq G(\eta^+_{j'},\ -\underline{w}_i\underline{\tau}^+,\ \eta^o_{j'},\ -\underline{w}_i\underline{\tau}^o)\ ,$$

where $G(a,b) = g[\Phi(a),\Phi(b)] + g[1-\Phi(a),\ 1-\Phi(b)]$. Furthermore, $G(a,b) \leq 0$ with equality only when $a=b$ (see e.g., Rao, 1973, pg. 58).

We now show that there exist positive constants P and δ_1 and integers n_o and j' such that for all $n > n_o$,

$$|(\eta^+_{j'},\ -\underline{w}_i\underline{\tau}^+) - (\eta^o_{j'},\ -\underline{w}_i\underline{\tau}^o)| \geq \delta_1 \tag{1.25}$$

holds at least nP times $(i=1,\ldots,n)$. This will establish (1.22) since, with $G(\cdot,\cdot)$ continuous and $|\eta^o_{j'}, -\underline{w}_i\underline{\tau}^0|$ bounded, there exists a $\delta_2 > 0$ such that $G(\eta^+_{j'}, -\underline{w}_i\underline{\tau}^+,\ \eta^o_{j'}, -\underline{w}_i\underline{\tau}^o) < -\delta_2$ whenever (1.25) holds.

Take j' to be such that $|\eta_{j'}^{+}-\eta_{j'}^{o}| \geq |\eta_{j}^{+}-\eta_{j}^{o}|$ (j=1,...,m-1) and define the vector $\underline{c}$ by $\underline{c} = (\eta_{j'}^{+}-\eta_{j'}^{o}, -\underline{\tau}^{+'}+\underline{\tau}^{o'})'$, so that

$$(1,\underline{w}_i)\underline{c} = (\eta_{j}^{+},-\underline{w}_i\underline{\tau}^{+}) - (\eta_{j}^{o},-\underline{w}_i\underline{\tau}^{o}).$$

From (1.21) we have that

$$\lim_{n\to\infty} n^{-1}\, \underline{c}'(\underline{1}_n,\underline{W}_n)'(\underline{1}_n,\underline{W}_n)\underline{c} = \underline{c}'\underline{D}\underline{c} \geq \lambda||\underline{c}||^2 > 0\ , \tag{1.26}$$

where λ = the smallest eigenvalue of $\underline{D}$ and $||\underline{a}||$ is the Euclidean norm of any vector $\underline{a}$. Inequality (1.26) implies the existence of an integer n_o such that for all $n > n_o$,

$$n^{-1}\sum_{i=1}^{n} [(1,\underline{w}_i)\underline{c}]^2 \geq \tfrac{1}{2}\lambda\ ||\underline{c}||^2\ . \tag{1.27}$$

Now $|(1,\underline{w}_i)\underline{c}|$ is bounded by $||\underline{c}||M_1$, where $M_1 = \sup_i ||(1,\underline{w}_i)||$ So (1.27) implies that at least nP of the n terms $[(1,\underline{w}_i)\underline{c}]^2$ are greater than δ_1^2, where

$$P = [\lambda/4(M_1^2 - \lambda/4)]$$

and

$$\delta_1 = \tfrac{1}{2}\lambda^{\frac{1}{2}}||c||\ .$$

This establishes (1.22) and thus Hoadley's condition C4(i). Hence, we have established that

$$\underline{\bar{\omega}}_n \to_p \underline{\omega}^o. \tag{1.28}$$

By definition, (1.28) is equivalent to

$\lim_{n\to\infty} \Pr[\,||\underline{\bar{\omega}}_n - \underline{\omega}^o||_\infty \leq \epsilon] = 1$ for any $\epsilon > 0$, where

$||(a_1,\ldots,a_s)'||_\infty = \mathrm{Max}[|a_1|,\ldots,|a_s|]$. Take ε to be any positive constant such that

$$||\underline{\omega}^o||_\infty < K-\varepsilon. \tag{1.29}$$

Now since $\ln[\phi(\cdot)]$ is concave, $L_n(\underline{\omega})$ is also concave (Burridge, 1981). Hence, if $||\underline{\bar{\omega}}_n|| < K$, then $L(\underline{\bar{\omega}}_n)$ is a local maximum and thus $L_n(\underline{\bar{\omega}}_n) = \sup_{\omega\in\Omega} L_n(\underline{\omega})$ and, since $\underline{\hat{\omega}}_n$ is unique, $\underline{\bar{\omega}}_n = \underline{\hat{\omega}}_n$. Whenever $||\underline{\bar{\omega}}_n - \underline{\omega}^o||_\infty < \varepsilon$, (1.29) implies that $||\underline{\bar{\omega}}_n||_\infty \leq K$, so

$$\lim_{n\to\infty} \Pr[\underline{\bar{\omega}}_n = \underline{\hat{\omega}}_n] \geq \lim_{n\to\infty} \Pr[||\underline{\bar{\omega}}_n - \underline{\omega}^o||_\infty \leq \varepsilon] = 1,$$

i.e., $\underline{\bar{\omega}}_n - \underline{\hat{\omega}}_n \to_p \underline{0}$, and the proof is complete.

<u>Theorem 2</u>: Consider the sequence of alternative threshold models (1.20). Let $\underline{\omega}^o \varepsilon \Omega^o$, where $\Omega^o = \{\underline{\omega}: -\infty < \eta_1 < \ldots < \eta_{m-1} < \infty, \underline{\tau}\varepsilon\underline{R}^r\}$. Suppose $\underline{\hat{\omega}}_n \to_p \underline{\omega}^o$. Suppose further that there is an open neighborhood $N(\underline{\omega}^o)$ of $\underline{\omega}^o$, and constants M_1, M_2, and M_3 such that, for all points $\underline{\omega}^+ \varepsilon N(\underline{\omega}^o)$:

1. The limit $\underline{B}(\underline{\omega}^+) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \underline{B}_i(\underline{\omega}^+)$ exists and is positive definite.

2. The first-, second-, and third-order partial derivatives of $\ln[P^*_{ij}(\underline{\omega})]$ with respect to the elements of $\underline{\omega}$ evaluated at $\underline{\omega}^+$ are bounded by M_1, M_2, and M_3, respectively ($i=1,2,\ldots$; $j=1,\ldots,m$).

Then $n^{\frac{1}{2}}(\underline{\hat{\omega}}_n - \underline{\omega}^o)$ coverges in distribution to the multivariate normal distribution with null mean vector and covariance matrix $[\underline{B}(\underline{\omega}^o)]^{-1}$.

Theorem 2 follows immediately from Hoadley's Theorem 2 upon verifying that his conditions N1-N9 are satisfied. Our conditions 1 and 2 ensure that his conditions N4 and N7-N9 are satisfied for all $\underline{\omega}\varepsilon N(\underline{\omega}^o)$. (That these four conditions hold for all $\underline{\omega}\varepsilon\Omega^o$ is not required in Hoadley's proof.) That the remainder of his conditions are satisfied is readily verified.

Condition 2 of Theorem 2 holds when $\underline{\omega}^o\varepsilon\Omega^o$, the conditions of Theorem 1 are satisfied, and $\Phi(\cdot)$ is the logistic or standard normal CDF. This claim can be verified by examining

$$\partial \ln[P^*_{ij}(\underline{\omega})]/\partial\tau_u = [\phi(\eta_{j-1}-\underline{w}_i\underline{\tau}) - \phi(\eta_j-\underline{w}_i\underline{\tau})]w_{iu}/P^*_{ij}(\underline{\omega})$$

and the other first-, second-, and third-order partial derivatives of $\ln[P^*_{ij}(\underline{\omega})]$. Suppose $\underline{\omega}^o\varepsilon\Omega^o$, the conditions of 1 are satisfied, and $\Phi(\cdot)$ is the standard normal or logistic CDF. Then there exists a $\delta^+ > 0$ such that $\eta^o_j - \eta^o_{j-1} \geq 2\delta^+$ for $j = 2,\ldots,m-1$. Since the elements of $\underline{w}_i$ are bounded, there exists a constant M^+ and an open neighborhood $N(\underline{\omega}^o)$ of $\underline{\omega}^o$ such that, for all $\underline{\omega}^+\varepsilon N(\underline{\omega}^o)$,

$$\eta^+_j - \eta^+_{j-1} > \delta^+ \quad (j=2,\ldots,m-1)$$

and

$$|\eta^+_j - \underline{w}_i\underline{\tau}^+| < M^+ \quad (i=1,\ldots,;\ j=1,\ldots,m-1).$$

Hence, for all $\underline{\omega}^+\varepsilon N(\underline{\omega}^o)$,

$$|\partial \ln[P^*_{ij}(\underline{\omega})]/\partial\tau_u|_{\underline{\omega}=\underline{\omega}^+}| = |[\phi(\eta^+_{j-1}-\underline{w}_i\underline{\tau}^+) - \phi(\eta^+_j-\underline{w}_i\underline{\tau}^+)]w_{iu}/P^*_{ij}(\underline{\omega}^+)$$

$$\leq \phi(0)M/[\delta^+\phi(M^+ + \delta^+)], \qquad (1.30)$$

since $\phi(t) \leq \phi(0)$, $|w_{iu}| \leq M$, and $P^*_{ij}(\underline{\omega}^+) = \int_{\eta^+_{j-1}-\underline{w}_i\underline{\tau}^+}^{\eta^+_j-\underline{w}_i\underline{\tau}^+} \phi(t)dt > \delta^+\phi(M^*+\delta^+)$.

[To establish the bound in (1.30), the only property of $\phi(\cdot)$ used is that $\phi(t)$ is a decreasing function of $|t|$.] That the second- and third-order partial derivatives of $\ln[P^*_{ij}(\underline{\omega}^+)]$ are bounded for all $\underline{\omega}^+ \varepsilon N(\underline{\omega}^o)$ can be established in a similar manner. Thus, we conclude that, for $\phi(\cdot)$ given by (1.2) or (1.3), the conditions which ensure consistency of the ML estimator also ensure some of the conditions of Theorem 2.

4. CONFIDENCE BOUNDS FOR PROBABILITIES

We now use the asymptotic results given in Section 3.2 to devise an approximate confidence interval for each of the following probabilities: (1) the probability that a future individual belongs to one of the first j' categories, i.e.,

$$\sum_{j=1}^{j'} P^*_{0j}(\underline{\omega}) = \Phi(\eta_{j'} - \underline{w}_0\underline{\tau}) , \tag{1.31}$$

and (2) the probability that a future individual belongs to one of c consecutive categories (where neither Category 1 nor Category m is included), i.e., the probability

$$\sum_{j=j'+1}^{j'+c} P^*_{0j}(\underline{\omega}) = \Phi(\eta_{j'+c}-\underline{w}_0\underline{\tau}) - \Phi(\eta_{j'}-\underline{w}_0\underline{\tau}). \tag{1.32}$$

Note that, when c=1, (1.32) reduces to the probability that a future individual belongs to a single category, namely Category (j'+1).

The confidence interval procedure involves two steps. In Step 1, we construct an approximate confidence interval for $\eta_{j'} - \underline{w}_0\underline{\tau}$ or approximate confidence region for $\eta_{j'+c} - \underline{w}_0\underline{\tau}$ and $\eta_{j'} - \underline{w}_0\underline{\tau}$ by acting as though the ML estimator $\hat{\underline{\omega}}$ or $\underline{\omega}$ is multivariate normal with variance-covariance matrix $[\underline{B}(\hat{\underline{\omega}})]^{-1}$. In Step 2, we obtain an approximate confidence interval for the probability (1.31) or (1.32) by locating the maximum and minimum value of the probability over the interval or region constructed in Step 1.

Let V_1 and V_2 represent bivariate normal observable random variables with unknown means μ_1 and μ_2, satisfying $\mu_2 \geq \mu_1$, and known covariance matrix $\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. Take C_δ to be the upper δ point of the standard normal distribution and let $\underline{E}_j$ represent the 1 x (m-1) unit vector with j^{th} element 1.

To obtain an approximate (1-δ)100% confidence interval for (1.31), we construct a (1-δ)100% confidence interval (L_p, U_p) for $\Phi(\mu_1)$ and then substitute $\hat{\eta}_{j'} - \underline{w}_0\hat{\underline{\tau}}$ and its asymptotic variance $(\underline{E}_{j'}, -\underline{w}_0)[\underline{B}(\hat{\underline{\omega}})]^{-1}(\underline{E}_{j'}, -\underline{w}_0)'$ for V_1 and σ_1^2, respectively. [(L_p, U_p) is a function of V_1 and σ_1^2]. Similarly, to obtain an approximate (1-δ)100% confidence interval for (1.32), we construct a (1-δ)100% confidence interval for $\Phi(\mu_2) - \Phi(\mu_1)$ and then substitute $\hat{\eta}_{j'} - \underline{w}_0\hat{\underline{\tau}}$ and $\hat{\eta}_{j'+c} - \underline{w}_0\hat{\underline{\tau}}$ and their asymptotic variances and covariance for V_1, V_2, σ_1^2, σ_2^2, and σ_{12}, respectively. We consider the construction of the confidence intervals, referring to the construction of the interval for $\Phi(\mu_1)$ as Case 1 and to that of the interval for

$\Phi(\mu_2) - \Phi(\mu_1)$ as Case 2.

<u>Case 1</u>: Let $P = \Phi(\mu_1)$. A $(1-\delta)100\%$ confidence interval for μ_1 is given by

$$(L_1, U_1) = (V_1 - C_{\delta_2}\sigma_1,\ V_1 + C_{\delta_1}\sigma_1)\ ,$$

where δ_1 and δ_2 are nonnegative constants such that $\delta_1 + \delta_2 = \delta$. Since $\Phi(\cdot)$ is a monotone increasing function,

$$(L_p, U_p) = [\Phi(L_1), \Phi(U_1)]$$

is a $(1-\delta)100\%$ confidence interval for P. For a two-sided interval for P, we can, e.g., take $\delta_1 = \delta_2 = \delta/2$. For a lower (an upper) confidence bound for P, take $\delta_1 = 0$ and $\delta_2 = \delta$ ($\delta_1 = \delta$ and $\delta_2 = 0$).

<u>Case 2</u>. Let $P_d = \Phi(\mu_2) - \Phi(\mu_1)$. Suppose A is a $(1-\delta)100\%$ confidence region for (μ_1, μ_2). If A contains points that violate the constraint $\mu_2 \geq \mu_1$, a smaller $(1-\delta)100\%$ confidence region B can be obtained by simply deleting such points. By determining the infinum L_p and supremum U_p of P_d over B, a confidence interval is obtained for P_d for which the confidence coefficient is at least $(1-\delta)$. (Note that, for some data sets, every point may violate the constraint $\mu_2 \geq \mu_1$ and the result of the deletion is an empty set. If B is empty, we arbitrarily take $L_p = U_p = 0$.)

Alternatively, we may obtain (L_p, U_p) directly from A (unless A is a disconnected set). First, determine the infinum L and supremum U of P_d over A, and then set

$$(L_p, U_p) = [\mathrm{Max}(0,L),\ \mathrm{Max}(0,U)] \tag{1.33}$$

For a given value of μ_2, P_d is a monotone decreasing function of μ_1. Hence, L and U (or L_p and U_p) must occur at points on the "boundary" of A (or B). (If the region is not bounded, the infinum and/or supremum of P_d may occur where μ_1 or μ_2 equals $-\infty$ or $+\infty$.)

There are a number of possible choices for the region A. The elliptical confidence region for μ_1 and μ_2 described, e.g., by Anderson (1958, pg. 55) could be used. It can be shown that locating the point at which P_d is maximized (or minimized) over such a region would require the solution of a nonlinear equation.

Alternatively, we could use the Bonferroni inequality (Miller, 1966, pg. 101) to obtain a rectangular region $\{(\mu_1,\mu_2):L_1 \le \mu_1 \le U_1;\ L_2 \le \mu_2 \le U_2\}$, where $L_i = V_i - C_{\delta_i}\sigma_i$ and $U_i = V_i + C_{\delta_i}\sigma_i$ (i=1,2), with confidence coefficient at least $(1-2\delta_1)(1-2\delta_2)$. For such a region,

$$L_p = \mathrm{Max}[\Phi(L_2) - \Phi(U_1),\ 0]$$

and

$$U_p = \mathrm{Max}[\Phi(U_2) - \Phi(L_1),\ 0]\ . \tag{1.34}$$

If the Bonferroni inequality is to be used when the correlation between V_1 and V_2 is large, it may be preferable to construct a rectangular region for μ_2 and $\mu_2 - \mu_1$, rather than for μ_1 and μ_2.

One-sided confidence bounds for P_d may also be constructed using the Bonferroni inequality. For example, an upper bound for P_d with a confidence coefficient of at least $(1-\delta)$ is given by U_p in (1.34), where $(1-\delta) = (1-\delta_1)(1-\delta_2)$.

We now consider a third possible choice for the confidence region A. This region consists of the points enclosed by a parallelogram. Unlike the elliptical region, this region is such that, in the case where Φ is the normal CDF, the points where P_d attains its maximum and minimum can be determined analytically. This region is obtained by forming a rectangular confidence region for μ_2 and a linear combination of μ_1 and μ_2 (chosen so that the corresponding linear combination of V_1 and V_2 is uncorrelated with V_2). This region may tend to produce tighter bounds for P_d than would a rectangular region for μ_1 and μ_2 (or for μ_2 and $\mu_2 - \mu_1$) since the confidence coefficient for the region is exact. The region for this approach is given in the following lemma:

<u>Lemma 2</u>: A $(1-\delta)100\%$ confidence region for (μ_1,μ_2) is defined by the two inequalities

$$V_1 - \lambda(V_2-\mu_2) - C_{\delta_1}\sigma_{1.2} \leq \mu_1 \leq V_1 - \lambda(V_2-\mu_2) + C_{\delta_2}\sigma_{1.2} \tag{1.35}$$

and

$$V_2 - C_{\delta_3}\sigma_2 \leq \mu_2 \leq V_2 + C_{\delta_4}\sigma_2 \ . \tag{1.36}$$

Here, $\lambda = \sigma_{12}/\sigma_2^2$, $\sigma_{1.2} = (\sigma_1^2 - \sigma_{12}^2/\sigma_2^2)^{\frac{1}{2}}$ and δ_1, δ_2, δ_3, δ_4 are non-negative constants such that $(1-\delta_1-\delta_2)(1-\delta_3-\delta_4) = (1-\delta)$.

<u>Proof</u>: The inequality (1.35) is a function of $(V_1-\lambda V_2)$ and the inequality (1.36) is a function of V_2. Since V_2 and $(V_1-\lambda V_2)$ are distributed independently, the confidence coefficient for the region defined by (1.35) and (1.36) equals $(1-\delta_1-\delta_2)(1-\delta_3-\delta_4)$, and thus

equals $(1-\delta)$.

In Figure 1, we illustrate the shape of the parallelogram region. Whenever $\ln[\phi(\cdot)]$ is concave, the contours for P_d will be as indicated by the dashed lines in the Figure. (The upper left contours correspond to higher values for P_d.) Since contours above $\mu_1=\mu_2$ have increasing slope, if the minimum for P_d over A is positive, it will necessarily occur at either the lower right or upper right corner of the parallelogram. Thus, L_p may be easily determined. The maximum will occur somewhere along the left boundary. We now indicate how U may be determined. (If $\ln[\phi(\cdot)]$ were not concave, L could be determined in a manner analogous to that of determining U.)

Denote the interval (1.36) for μ_2 by $[L_2,U_2]$ and the left side of (1.35) by $L_1(\mu_2)$. For a given value of μ_2 the maximum value of P_d is

$$\Phi(\mu_2) - \Phi[L_1(\mu_2)]. \tag{1.37}$$

To determine U, it remains to maximize (1.37) for $\mu_2 \varepsilon [L_2,U_2]$. To do so, we equate the derivative of (1.37) (with respect to μ_2) to zero, obtaining

$$\phi(\mu_2) - \lambda\phi[L_1(\mu_2)] = 0. \tag{1.38}$$

By using (1.38), we can determine whether (1.37) has any stationary points. If $\lambda \leq 0$, (1.38) has no solution. If $\lambda > 0$, we can attempt to solve (1.38) for μ_2. When ϕ is the standard normal density, (1.38) is equivalent to

$$\mu_2^2 = [L_1(\mu_2)]^2 - 2\ \ln\lambda. \tag{1.39}$$

Substituting $L_1(\mu_2) = V_1 - \lambda(V_2-\mu_2) - C_{\delta_1}\sigma_{1.2}$ in (1.39), we obtain a quadratic equation in μ_2, with (real or complex) solutions

$$\{\lambda(V_1-\lambda V_2-C_{\delta_1}\sigma_{1.2}) \pm [(V_1-\lambda V_2-C_{\delta_1}\sigma_{1.2})^2-2(1-\lambda^2)\ln\lambda]^{\frac{1}{2}}\}/(1-\lambda^2) \tag{1.40}$$

Therefore, U can be computed as the maximum value of (1.37) over several points: L_2, U_2, and any real solution to the quadratic equation (1.39) that is contained in (L_2, U_2). Then U_p is obtained by (1.33).

5. EXAMPLE

Ashford (1959) analyzed survey data on the occurrence of pneumoconiosis among coal miners using the threshold model. He took Φ to be the normal CDF. We reconsider this application, using it to illustrate our confidence-interval procedures.

Pneumoconiosis is a lung disease caused by the continual inhalation of irritating particles. The survey classified each miner into one of 8 groups based on the number of years of mining and into one of 3 groups based on the degree of abnormality revealed by an X-ray. Ashford states, "This classification (by degree of abnormality) corresponds to an arbitrary subdivision of the continuous scale of abnormality associated with simple pneumoconiosis into three ordered and mutually exclusive classes." Apparently, Category 1 corresponded to a normal X-ray, Category 2 to the first stage of simple pneumoconiosis, and Category 3 to the more severe stages of pneumoconiosis [see Fay (1957)]. The data given in Table 1 are those from a sample of 371 men who had worked primarily in the same type of mining. The number listed under 'Years' is the midpoint of the group.

These data were taken from the early part of a long-term field study conducted by United Kingdom National Coal Board. The purpose of this study was to determine the effect of the dust inhaled by coal miners in their work. The researchers hoped "to obtain accurate data on which to base safe levels of dust concentrations, which miners will be able to tolerate throughout their working lives without suffering any considerable difficulty" (Fay, 1957, pg. 309). A conceivable further objective in the analysis of such data might be the determination of when (in terms of years on occupation) miners should be checked by X-ray for the presence of pneumoconiosis. In making such a decision, the risks from the disease would have to be balanced against the hazards of X-rays.

We apply the threshold model to the data in Table 1 to estimate the prevalence of pneumoconiosis among miners with given lengths of exposure. Categories 2 and 3 combined represent those men having abnormal X-rays. Computing upper confidence bounds for the frequency of these two categories among miners with a given number of years of exposure might be helpful in deciding when to administer chest X-rays. We illustrate the confidence interval procedure by obtaining an upper confidence bound for the frequency of abnormal X-rays among coal miners with 25 years of exposure. To further illustrate, we construct confidence intervals for the frequency of Category 2 among coal miners with 25 years of exposure and the frequency of Category 3.

Following Ashford, we apply a threshold model in which

$$y_i = \alpha_1 + \alpha_2 x_i + e_i ,$$

where the underlying conceptual variable y_i reflects the extent of abnormality, x_i denotes log (to the base 10) of years in occupation for the i^{th} miner, and α_j is an unknown parameter ($j = 1,2$). The quantities ξ_1 and ξ_2 represent unknown boundary points which define the three intervals on the y-scale.

The parameters ξ_1, ξ_2, α_1, α_2, and σ of the threshold model are not identifiable. Ashford dealt with the problem of nonidentifiability by setting $\xi_1 = 0$ and $\xi_2 = 1$. Estimates $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\sigma}$ obtained by Ashford via ML represent ML estimates of the identifiable functions $(\alpha_1-\xi_1)/(\xi_2-\xi_1)$, $\alpha_2/(\xi_2-\xi_1)$, and $\sigma/(\xi_2-\xi_1)$ respectively. [When $\underline{1}\varepsilon M(\underline{X})$ and each category contains observations, any two boundary points can be chosen "arbitrarily" without placing any constraints on identifiable functions of $\underline{\theta}$.]

We carried out the ML computation by applying the Newton-Raphson procedure to the alternative model. The ML estimates of the parameters η_1, η_2, and τ_1 of the alternative model and their estimated variances and covariances were:

$$\begin{aligned} \hat{\eta}_1 &= 5.460, \\ \hat{\eta}_2 &= 5.986, \\ \hat{\tau}_1 &= 3.359, \end{aligned} \tag{1.41}$$

and

$$\text{Cov}(\hat{\eta}_1, \hat{\eta}_2, \hat{\tau}_1) = \begin{bmatrix} .50985 & .51386 & .34023 \\ .51386 & .52409 & .34428 \\ .34023 & .34428 & .23022 \end{bmatrix}.$$

The estimated P_{ij}'s are given in Table 2. The Pearson χ^2 statistic for lack of fit was 3.516 on 13 d.f. (P = .995), indicating that the model provides a good fit to the data. (Any inaccuracy in the χ^2 approximation due to small expected frequencies in the 5.8-year group would not affect this inference.)

Suppose we wish to assess the chances that a miner with 25 years exposure will belong to a particular category. With $w_0 = \log_{10}(25) = 1.398$, the estimated probabilities are:

$$P^*_{01}(\underline{\hat{\omega}}) = \Phi(\hat{\eta}_1 - 1.398\ \hat{\tau}_1) = \Phi(.764) = .777,$$

$$P^*_{02}(\underline{\hat{\omega}}) = \Phi(\hat{\eta}_2 - 1.398\ \hat{\tau}_1) - P^*_{01}(\underline{\hat{\omega}}) = \Phi(1.290) - .777 = 1.24,$$

and

$$P^*_{03}(\underline{\hat{\omega}}) = 1 - \Phi(\hat{\eta}_2 - 1.398\ \hat{\tau}_1) = 1 - \Phi(1.290) = .098$$

for Categories 1,2, and 3, respectively.

An approximate 90% lower confidence bound for $(\eta_1 - 1.398\tau_1)$ is .764-.118 = .646. Therefore, an approximate 90% upper confidence bound for $[P^*_{02}(\underline{\omega}) + P^*_{03}(\underline{\omega})]$ is $1 - \Phi(.646) = .259$.

An approximate 90% confidence interval for $(\eta_2 - 1.398\tau_1)$ is

$$(1.29-.176,\ 1.29+.176) = (1.114,\ 1.466).$$

The corresponding approximate 90% confidence interval for $P^*_{03}(\underline{\omega})$ is (.071, .133).

An approximate 90% confidence region for $(\eta_1-1.398\tau_1)$ and $(\eta_2-1.398\tau_1)$, using the parallelogram method described in Section 4 under Case 2, is determined by the two inequalities

$$1.082 \le (\eta_2-1.398\tau_1) \le 1.499$$

and

$$-.089 + .6(\eta_2-1.398\tau_1) \le (\eta_1-1.398\tau_1) \le .069 + .6(\eta_2-1.398\tau_1).$$

The maximum value of $P^*_{02}(\underline{\omega})$ over this confidence region occurs on the left boundary. Using (1.40), we find that (1.37) has a single stationary point for $\mu_2 = \eta_2-1.398\tau_1 \in [1.082, 1.499]$. It occurs at $\mu_2 = 1.188$. The value of $P^*_{02}(\underline{\omega})$ at this point is

$$\Phi(1.188) - \Phi[-.089 + .6(1.188)] = .149.$$

Since this value is larger than the value at either endpoint of the left boundary, the maximum value for $P^*_{02}(\underline{\omega})$ over the entire confidence region is .149.

The minimum value for $P^*_{02}(\underline{\omega})$ in this region occurs at the lower endpoint of the right boundary and is .097. Therefore, an approximate 90% confidence interval for $P^*_{02}(\underline{\omega})$ is (.097, .149).

Note that the correlation between pairs of estimates is nearly one. When the probability of a response in the second category appears to be small, we could anticipate that the estimates of η_1 and η_2 will tend to be close together and thus highly correlated. The parameters η_j and τ_1 correspond to the parametric functions $(\xi_j-\alpha_1)/\sigma$ and α_2/σ, respectively. To suggest why $\hat{\eta}_j$ and $\hat{\tau}_1$ are highly correlated, note that, if the underlying responses $y_1,\ldots,y_n$ could be observed, the correlation between the ML estimates for α_1 and α_2 would be -.967. Thus, the large estimates (.993 and .991) obtained for $\mathrm{Corr}(\hat{\eta}_1,\hat{\tau}_1)$ and $\mathrm{Corr}(\hat{\eta}_2,\hat{\tau}_1)$ are not primarily a consequence of the categorical nature of the data.

Because the 'x-values' that are available from the study are the midpoints of intervals rather than actual number of years of exposure, there is an errors-in-variables aspect to the problem of estimating the probability that a miner with a given number of years of exposure will belong to a particular category. In illustrating our procedures, we have chosen to ignore this aspect. It should also be noted that miners who develop pneumoconiosis may tend to change occupations, in which case they may be under-represented in the data. Such a tendency could affect the appropriateness of the model and introduce some bias into the analysis.

Our estimates (1.41) of the parameters of the alternative threshold model differ somewhat from those constructed from Ashford's results. Ashford obtained $\hat{\alpha}_1 = -7.52$, $\hat{\alpha}_2 = 4.61$, and $\sigma = 1.56$, which, e.g., gives $\hat{\eta}_1 = 4.82$ rather than $\hat{\eta}_1 = 5.46$. However, the difference between the value of the likelihood function corresponding to Ashford's estimates and that corresponding to the estimates (1.41) is small. As suggested by the "near singularity" of the information matrix, the likelihood function seems to have a nearly stationary ridge through the maximum.

Table 1. Period of Exposure and Prevalence of Pneumoconiosis among a Sample of Coal Miners (Ashford, 1959)

Years	Degree of Abnormality		
	1	2	3
5.8	98	0	0
15.0	51	2	1
21.5	34	6	3
27.5	35	5	8
33.5	32	10	9
39.5	23	7	8
46.0	12	6	10
51.5	4	2	5

Table 2. Estimated Probabilities

Years	Degree of Abnormality 1	2	3
5.8	.998	.002	.000
15.0	.934	.045	.021
21.5	.838	.097	.065
27.5	.734	.141	.125
33.5	.632	.174	.194
39.5	.539	.195	.266
46.0	.450	.206	.344
51.5	.386	.207	.407

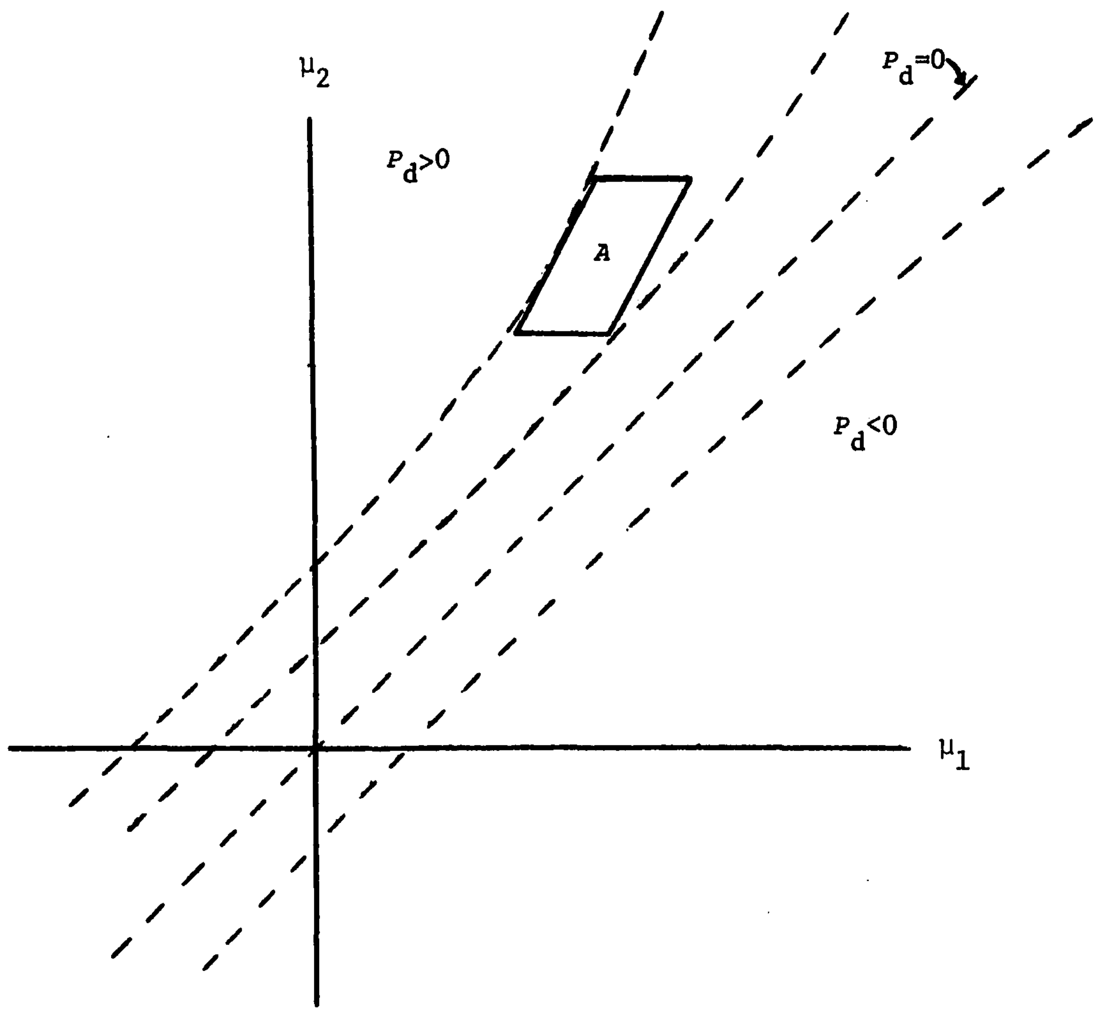

Figure 1. Confidence Region for (μ_1,μ_2) and Contours for P_d

6. REFERENCES

Aitchison, J. and Silvey, S.D. 1957. The generalization of probit analysis to the case of multiple responses. Biometrika 44, 131-140.

Anderson, J.A. 1981. Discussion of: Regression models for ordinal data. Journal of the Royal Statistical Society Ser. B 42, 130-131.

Anderson, T.W. 1958. An Introduction to Multivariate Statistical Analysis. New York: Wiley.

Ashford, J.R. 1959. An approach to the analysis of data for semi-quantal responses in biological assay. Biometrics 15, 573-581.

Bock, R. D. 1975. Multivariate Statistical Methods in Behavioral Research. New York: McGraw Hill, Inc.

Burridge, J. 1981. A note on maximum likelihood estimation for regression models using grouped data. Journal of the Royal Statistical Society Ser. B 43, 41-45.

Fay, J.W.J. 1957. The national coal board's penumoconiosis field research. Nature 180, 309-11.

Finney, D.J. 1971. Probit Analysis, 3rd Ed, Cambridge: Cambridge University Press.

Gurland, J., Lee, I. and Dahm, P.A. 1960. Polychotomous quantal response in biological assay. Biometrics 16,382-398.

Haberman, S.J. 1980. Discussion of: Regression models for ordinal data. Journal of the Royal Statistical Society Ser. B 42, 136-137.

Hoadley, B. 1971. Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. The Annals of Mathematical Statistics 42, 1977-1991.

McCullagh, P. 1980. Regression models for ordinal data. Journal of the Royal Statistical Society Ser. B 42, 109-127.

McKelvey, R.D. and Zavoina, W. 1975. A statistical model for the analysis of ordinal level dependent variables. Journal of Mathematical Sociology 4, 103-120.

Miller, R.G. Jr. 1966. Simultaneous Statistical Inference. New York: McGraw Hill, Inc.

Rao, C.R. 1973. Linear Statistical Inference and Its Applications. New York: Wiley.

Snell, E.J. 1964. A scaling procedure for ordered categorical data. Biometrics 20, 592-607.

Williams, O.D. and Grizzle, J.E. 1972. Analysis of contingency tables having ordered response categories. Journal of the American Statistical Association 67, 55-63.

Wright, S. 1934. The results of crosses between inbred strains of guinea pigs, differing in number of digits. Genetics 19, 537-551.

PART II.

ANALYSIS OF ORDINAL DATA

VIA THE THRESHOLD MODEL

SUMMARY

The response variable in a biological study is often an ordered categorical variable. A possible model for such data is the threshold model, which is defined in terms of an underlying continuous random variable that follows a linear model. An iterative computational procedure is required to apply maximum likelihood. Among the procedures that have been proposed are the Newton-Raphson and method-of-scoring procedures. We introduce a third iterative procedure that may be more economical for large data sets. We provide several examples to illustrate the suitability of the threshold approach, to com-evaluate the performance of the computational algorithms, and to compare the logistic and normal versions of the threshold model.

Key words: Ordered categories; maximum likelihood; EM algorithm; logistic distribution; multiple regression.

1. INTRODUCTION

Frequently, the response variable which is to assume the role of the dependent variable in a statistical analysis of biological or agricultural data is (for reasons of convenience or necessity) not measured precisely. For example, beef cattle are graded standard, good, choice, or prime. Here, the categories (grades) consist of individuals (cattle) which are of similar but not identical "quality". Also the categories have a natural ordering with respect to quality. Examples with similar characteristics are calving difficulty (Berger and Freeman, 1978), severity of lung disease (Ashford, 1959), damage to a corn plant by pests (Guthrie, et al., 1978), and insect development (Aitchison and Silvey, 1957).

Many statistical procedures for analyzing ordered categorical responses have appeared in the literature [e.g., Aitchison and Silvey (1957), Ashford (1959), Williams and Grizzle (1972), and Andrich (1979)] and have been tried by practitioners. Most commonly, perhaps, successive integers are assigned to the ordered categories and a linear model is fitted to these integers. Although such an approach seems quite natural, it can result in potentially severe biases (see, e.g., McKelvey and Zavoina, 1975).

Here, we consider the approach presented by Ashford (1959) and discussed by Bock (1975), McKelvey and Zavoina (1975), and McCullagh (1980). The model is based on the assumption that the observed category is determined by the value of an underlying continuous random variable for which a linear model is applicable. As described by Bock (1975, pg. 513),

> "The underlying process is assumed scalar valued and distributed continuously in the population of subjects. There are assumed to be certain values on the continuum called thresholds, such that the m response categories correspond to intervals from $-\infty$ to $+\infty$ defined by the m-1 threshold values. The response of a given subject is determined by the interval in which his process value falls...."

We will refer to this approach as the threshold approach. When the original responses are actually continuous (but for means of convenience only categorical information is recorded) or when continuous responses can be conceptualized, the threshold approach is appealing. In Section 2 we give a precise definition of the threshold model.

Maximum likelihood (ML) has generally been used to estimate the parameters of the threshold model. An iterative procedure is required for the computations. Ashford used the Newton-Raphson (NR) procedure, as did McKelvey and Zavoina. Both Bock and McCullagh used the method of scoring (MS). Snell (1964) indicated that he had used the method of steepest ascent.

Two objectives of this paper are to introduce a new algorithm and to give general considerations in deciding on an algorithm. Another purpose is to illustrate the application of the threshold model to biological settings and to exhibit the usefulness of the threshold model approach relative to other approaches. The estimates of the threshold model parameters and (generally to a lesser extent) the conclusions of the analysis are dependent on the choice of the underlying distribution. We will consider the importance of choosing between the normal and logistic distributions. A final objective is to compare the actual performance of algorithms.

In Section 3.1, we discuss the implementation of the NR and the MS procedures. In Section 3.2, we introduce an alternative iterative procedure that may be more economical for large data sets than either NS or MS.

In Section 4, we illustrate the use of the threshold model in terms of several examples. The first three of these are taken from papers which propose alternative methods of analysis. We contrast the analyses based on the threshold model with the analyses produced by these alternative methods. The fourth example is one in which the number of categories and the number of parameters in the linear model are large and the data relatively sparse. Its inclusion was stimulated by questions raised by Anderson (1980) concerning the applicability of the threshold model under such circumstances.

In Section 5, we compare the performance of NR, MS, and the procedure described in Section 3.2 for the examples considered in Section 4.

2. THRESHOLD MODEL

Suppose that the response for the i^{th} of n "individuals" or "items" falls into one of m ordered categories. We number the categories 1,2,...,m, and take Z_i to be the discrete random variable defined by $Z_i = j$ when the i^{th} individual belongs to Category j. The observable response Z_i is assumed to be related to an unobservable, underlying continuous response variable t_i as follows:

$$\eta_{j-1} < t_i \leq \eta_j \iff Z_i = j,$$

where $j \in \{1,2,\ldots,m\}$, $\eta_0 = -\infty$, $\eta_m = +\infty$, and $\eta_1,\ldots,\eta_{m-1}$ are unknown boundaries (thresholds) which define a partitioning of the real line into m intervals.

Suppose that the underlying continuous response variable t_i satisfies the linear model

$$t_i = \underline{w}_i\underline{\tau} + e_i \qquad (i = 1,\ldots,n) \tag{2.1}$$

where

$\underline{w}_i = (w_{i1},\ldots,w_{ir})$ is a known vector, whose elements represent the values of r "independent" variables for the i^{th} individual,

$\underline{\tau} = (\tau_1,\ldots,\tau_r)'$ is a vector of unknown parameters or "fixed effects",

and

e_i = a random error.

We assume that $e_1,\ldots,e_n$ are independently and identically distributed with mean zero. We denote by $\Phi(\cdot)$ and $\phi(\cdot)$ the cumulative distribution function (CDF) and probability density function, respectively, of e_i, which we assume to be known. Common choices for Φ are the standardized normal CDF,

$$\Phi(t) = \int_{-\infty}^{t} (2\pi)^{-\frac{1}{2}} \; e^{-\frac{1}{2}u^2} du \; ,$$

and the logistic CDF,

$$\Phi(t) = (1-e^{-t})^{-1}.$$

[Since the boundaries are taken to be unknown, the variance of e_i is "nonestimable". Therefore specifying this variance places no restrictions on estimable functions. The details are presented by Mee and Harville (1981).]

Let

$$\underline{W} = \begin{bmatrix} \underline{w}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underline{w}_n \end{bmatrix}$$

and let $\underline{1}$ = a column vector of 1's. We assume that the matrix $(\underline{1},\underline{W})$ has full column rank. If this assumption were not satisfied, we could reparameterize the model in such a way that it would be satisfied by the reparameterized model. To illustrate, suppose we have a 2-way classification fixed-effects model (without interaction), i.e.,

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \quad (i=1,\ldots,I,\ j=1,\ldots,J).$$

Take the matrix $(\underline{1},\underline{W})$ to be the model matrix corresponding to the full rank reparameterization obtained by substituting $-(\alpha_1+\ldots+\alpha_{I-1})$ and $-(\beta_1+\ldots+\beta_{J-1})$ for α_I and β_J, respectively. For example, with I=2 and J=3,

$$\underline{W} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} \qquad (2.2)$$

Let $\underline{\omega} = (\eta_1,\ldots,\eta_{m-1}, \underline{\tau}')'$ and take the parameter space for $\underline{\omega}$ to be

$$\Omega = \{\underline{\omega}: \eta_0 \leq \eta_1 \leq \ldots \leq \eta_{m-1} \leq \eta_m,\ \underline{\tau} \in \underline{R}^r\}.$$

We refer to the above model for the observable responses as the threshold model.

Under the threshold model, the probability, P_{ij} that the i^{th} individual belongs to Category j is given by

$$P_{ij} = \Pr(\eta_{j-1} < t_i \leq \eta_j) = \Phi(\eta_j - \underline{w}_i\underline{\tau}) - \Phi(\eta_{j-1} - \underline{w}_i\underline{\tau}).$$

Let $\underline{Z} = (Z_1,\ldots,Z_n)'$. Note that $Z_1,\ldots,Z_n$ are independently distributed due to the assumed independence of the e_i's. Hence, the joint probability mass function $P(\cdot|\underline{\omega})$ of $\underline{Z}$ is

$$P[(z_1,\ldots,z_n)'|\underline{\omega}] = \prod_{i=1}^{n} P_{iz_i} \qquad (z_1=1,\ldots,m;\ldots;z_n=1,\ldots,m).$$

3. MAXIMUM LIKELIHOOD ESTIMATION

The likelihood function (ℓ) for the threshold model can be expressed as

$$\ell(\underline{\omega}) = \prod_{j \in C} \prod_{i \in R_j} P_{ij},$$

where R_j = {i:observed category for the i^{th} individual is Category j; i=1,...,n} (j=1,...,m) and C = {j:$R_j \neq$ empty set}. The ML estimate of $\underline{\omega}$ (if it exists) is defined to be an element $\hat{\underline{\omega}} = (\hat{\eta}_1,\ldots,\hat{\eta}_{m-1}, \hat{\underline{\tau}}')'$ of Ω such that

$$\ell(\hat{\underline{\omega}}) = \sup_{\underline{\omega} \in \Omega} \ell(\underline{\omega}) .$$

In practice, $\underline{\omega}$ is obtained by maximizing the log-likelihood function (L), where

$$L(\underline{\omega}) = \sum_{j \in C} \sum_{i \in R_j} \ln(P_{ij}) .$$

Haberman (1980) gave necessary and sufficient conditions for the existence of ML estimates for the threshold model when the parameter space is taken to be $\Omega^o = \{\underline{\omega}:\eta_0 < \eta_1 < \ldots < \eta_m, \underline{\tau} \in \underline{R}^r\}$. When the parameter space is taken to be Ω, i.e., when points with $\eta_j = \eta_{j-1}$ are included, his condition that every category contain at least one observation is not necessary. Let n_j = the number of observations in the j^{th} category. If, for some j, $n_j = 0$, then $\hat{\eta}_j = \hat{\eta}_{j-1}$ (Mee and Harville, 1981). Therefore, for purposes of computing the ML estimates, we may proceed by deleting any empty categories, reducing m accordingly, and renumbering the categories. Subsequently, we assume that $n_j > 0$ (j=1,...,m).

3.1 Iterative Procedures for Maximizing L

Beginning with an initial value $\underline{\omega}^{(0)}$ for $\underline{\omega}$, we obtain a sequence of values

$$\underline{\omega}^{(k+1)} = \underline{\omega}^{(k)} + \rho^{(k)}\underline{\nu}^{(k)} \qquad k = 0,1,\ldots, \tag{2.3}$$

such that the sequence $\{L[\underline{\omega}^{(k)}]\}$ is strictly increasing. The vector $\underline{\nu}^{(k)}$ in (2.3) determines the direction of change in obtaining the new value $\underline{\omega}^{(k+1)}$, while the scalar $\rho^{(k)}$ allows for varying the step length. The choice of $\underline{\nu}^{(k)}$ and $\rho^{(k)}$ distinguish one algorithm from another.

Let $\underline{\omega}^{(k)} = [\eta_1^{(k)},\ldots,\eta_{m-1}^{(k)}, \underline{\tau}^{(k)'}]'$. For the initial value $\underline{\omega}^{(0)}$ we choose $\underline{\tau}^{(0)} = \underline{0}$, a vector of 0's, and take $\eta_j^{(0)}$ (j-1,...,m-1) to be the solution to

$$\Phi[\eta_j^{(0)}] = (n_1+\ldots+n_j)/n.$$

These initial values for the boundaries are their ML estimates with $\underline{\tau}$ is constrained to be the null vector.

In both the NR and MS procedures, $\underline{\nu}^{(k)}$ is taken to be a solution to a system of equations of the form

$$\underline{G}^{(k)}\underline{\nu}^{(k)} = \partial L/\partial\underline{\omega}\Big|_{\underline{\omega}=\underline{\omega}^{(k)}} \quad , \tag{2.4}$$

where $\underline{G}^{(k)}$ is an s x s matrix (s = m-1+r). In the NR procedure, $-\underline{G}^{(k)}$ is taken to be the matrix of second-order partial derivatives of L and, in the MS procedure, it is taken to be the matrix of expected values of the second-order partial derivatives, evaluated at $\underline{\omega} = \underline{\omega}^{(k)}$.

After determining $\underline{\nu}^{(k)}$, we choose $\rho^{(k)}$. We require that $\rho^{(k)}$ be sufficiently small to ensure that $\underline{\omega}^{(k+1)}$ is in the interior of Ω, i.e., that $\eta_j^{(k+1)} > \eta_{j-1}^{(k+1)}$ (j=1,...,m).

We propose the choice $\rho^* = \text{Minimum}\ (1, \frac{1}{2}c)$, where c is the largest number such that, for all $\rho \in [0,c]$, $[\underline{\omega}^{(k)} + \rho\underline{\nu}^{(k)}] \in \Omega$. With this choice we never step more than halfway to the boundary of Ω. [If, for some j, $\eta_j = \eta_{j-1}$, then $\ell(\underline{\omega}) = 0$. Hence, we should not step "too near" the boundary, since $L(\underline{\omega})$ will approach $-\infty$ as $\underline{\omega}$ nears the boundary of Ω.]

We now set $\underline{\omega}^* = \underline{\omega}^{(k)} + \rho^*\underline{\nu}^{(k)}$, compute $L(\underline{\omega}^*)$, and determine whether the procedure has converged by monitoring the change in L(or ℓ). If the improvement satisfies some prescribed criteria (e.g., % change of ℓ greater than S%, for a given value S), set $\underline{\omega}^{(k+1)} = \underline{\omega}^*$ and proceed, i.e., determine $\underline{\omega}^{(k+2)}$. If the improvement fails to satisfy the criteria, we evaluate $L(\underline{\omega})$ at another point in the direction $\underline{\nu}^{(k)}$ determined as follows: use $L[\underline{\omega}^{(k)}]$ and $L(\underline{\omega}^*)$, together with the directional derivative of $L(\underline{\omega})$ at $\underline{\omega} = \underline{\omega}^{(k)}$ in the direction $\underline{\nu}^{(k)}$ to determine a quadratic approximation to $L[\underline{\omega}^{(k)} + \rho\underline{\nu}^{(k)}]$; find the value ρ^o which maximizes this approximation; determine $\rho^+ = \text{Minimum}\ (\rho^o, \frac{1}{2}c)$ and compute $L(\underline{\omega}^+)$, where $\underline{\omega}^+ = \underline{\omega}^{(k)} + \rho^+\underline{\nu}^{(k)}$. Now set $\underline{\omega}^{(k+1)}$ equal to the "best" of $\underline{\omega}^{(k)}$, $\underline{\omega}^*$, and $\underline{\omega}^+$, i.e., the point at which $L(\underline{\omega})$ is largest. Now if $\underline{\omega}^{(k+1)}$ satisfies the prescribed criteria, determine $\underline{\omega}^{(k+2)}$. Otherwise, stop.

We now provide expressions for the partial derivatives required for the NR and MS procedures and then conclude this subsection by comparing the computations involved in computing the second derivatives and expected second derivatives of $L(\underline{\omega})$.

Assume that $\phi'(t) = d\phi(t)/dt$ exists for all t, and let ϕ_{ij} and ϕ'_{ij} denote $\phi(\eta_j - \underline{w}_i\underline{\tau})$ and $\phi'(\eta_j - \underline{w}_i\underline{\tau})$, respectively. The first-order partial derivatives of L with respect to the elements of $\underline{\omega}$ are

$$\partial L/\partial\tau_u = \sum_{j\,\varepsilon\, C}\ \sum_{i\,\varepsilon\, R_j} w_{iu}(\phi_{i,j-1} - \phi_{ij})/P_{ij} \qquad (u=1,\ldots,r)$$

and

$$\partial L/\partial\eta_v = \sum_{i\,\varepsilon\, R_v} \phi_{iv}/P_{iv} - \sum_{i\,\varepsilon\, R_{v+1}} \phi_{iv}/P_{i,v+1} \qquad (v=1,\ldots,m-1).$$

The second-order partial derivatives of L with respect to the elements of $\underline{\omega}$ are:

$$\partial^2 L/\partial\tau_u\partial\tau_v = -\sum_{j\varepsilon C}\ \sum_{i\varepsilon R_j} w_{iu}w_{iv}\{[(\phi_{i,j-1}-\phi_{ij})/P_{ij}]^2 + (\phi'_{i,j-1}-\phi'_{ij})/P_{ij}\} \qquad (u,v=1,\ldots,r),$$

$$\partial^2 L/\partial\tau_u\partial\eta_v = -\sum_{i\varepsilon R_v} w_{iu}[(\phi_{i,v-1}-\phi_{iv})\phi_{iv}/P^2_{iv} + \phi'_{iv}/P_{iv}]$$

$$+ \sum_{i\varepsilon R_{v+1}} w_{iu}[(\phi_{iv}-\phi_{i,v+1})\phi_{iv}/P^2_{i,v+1}+\phi'_{i,v}/P_{i,v+1}] \qquad (u=1,\ldots,r;v=1,\ldots,m-1),$$

and

$$\partial^2 L/\partial\eta_u\partial\eta_v = \begin{bmatrix} \sum_{i\varepsilon R_v}[\phi'_{iv}/P_{iv}-(\phi_{iv}/P_{iv})^2]-\sum_{i\varepsilon R_{v+1}}[\phi'_{iv}/P_{i,v+1}+(\phi_{iv}/P_{i,v+1})^2] \\ \qquad\qquad\qquad\qquad (u=v=1,\ldots,m-1), \\ \sum_{i\varepsilon R_u}\phi_{i,u-1}\phi_{iu}/P^2_{iu} \qquad (u=v+1=2,\ldots,m-1), \\ \sum_{i\varepsilon R_v}\phi_{i,v-1}\phi_{iv}/P^2_{iv} \qquad (u=v-1=1,\ldots,m-2), \\ 0 \qquad \text{otherwise.} \end{bmatrix}$$

The expected second-order partial derivatives of L with respect to the elements of $\underline{\omega}$ are:

$$E[\partial^2 L/\partial\tau_u \partial\tau_v] = -\sum_{j=1}^{m} \sum_{i=1}^{n} w_{iu} w_{iv} (\phi_{i,j-1} - \phi_{ij})^2 / P_{ij} \quad (u,v-1,\ldots,r),$$

$$E[\partial^2 L/\partial\tau_u \partial\eta_v] = -\sum_{i=1}^{n} w_{iu} \phi_{iv} [(\phi_{i,v-1} - \phi_{iv})/P_{iv} - (\phi_{iv} - \phi_{i,v+1})/P_{i,v+1}]$$
$$(u=1,\ldots,r;\ v=1,\ldots,m-1),$$

and

$$E[\partial^2 L/\partial\eta_u \partial\eta_v] = \begin{bmatrix} -\sum_{i=1}^{n} \phi_{iv}^2 (P_{iv}^{-1} + P_{i,v+1}^{-1}) & (u=v=1,\ldots,m-1) \\ \sum_{i=1}^{n} \phi_{i,v-1} \phi_{iv} / P_{iv} & (u=v-1=1,\ldots,m-2) \\ \sum_{i=1}^{n} \phi_{i,u-1} \phi_{iu} / P_{iu} & (u=v+1=2,\ldots,m-1) \\ 0 & \text{otherwise.} \end{bmatrix}$$

Note that the expected second derivatives (ESD) do not involve $\phi'(\cdot)$ as the second derivatives (SD) do. However, in general, the ESD will involve many more computations since the summations over i in the SD are restricted to i-values corresponding to individuals in a particular category. Hence, the cost of computing the SD is expected to be low relative to that of computing the EDS. Note, however, that this general conclusion does not apply if the number of distinct $\underline{w}_i$ is very small relative to n. If $\underline{w}_i = \underline{w}_{i'}$, then the terms in the ESD for i and i' are equal. By exploiting this property, the number of computations can be reduced. If $\underline{w}_i = \underline{w}_{i'}$ and $Z_i = Z_{i'}$, then the terms in the SD for i and i' are equal, which can likewise be used to advantage in computing the SD. These considerations may be important in deciding which algorithm to use to compute ML estimates.

3.2 Normal Equation Procedure

We now introduce a third procedure for determining $\underline{\nu}^{(k)}$ which, in some situations, will require substantially fewer computations per iteration than either MS or NR. This procedure may be motivated as follows.

Suppose t_i, the underlying continuous variable, is distributed normally with mean $\underline{w}_i\underline{\tau}$ and variance 1. Then

$$E(t_i|Z_i=j) = E[t_i|t_i \varepsilon (\eta_{j-1},\eta_j)] = \underline{w}_i\underline{\tau} + (\phi_{i,j-1}-\phi_{ij})/P_{ij}$$

(see, e.g., Johnson and Kotz, 1970, pg. 81). Making the substitution

$$(\phi_{i,j-1}-\phi_{ij})/P_{ij} = E(t_i|Z_i = j)-\underline{w}_i\underline{\tau}$$

in the likelihood equation $\partial L/\partial\tau_u = 0$, we obtain

$$\sum_{j \varepsilon C} \sum_{i \varepsilon R_j} w_{iu}[E(t_i|Z_i=j)-\underline{w}_i\underline{\tau}] = 0 \quad (u=1,\ldots,r). \tag{2.5}$$

Let z_i represent the observed value of Z_i, let $t_i^* = E(t_i|Z_i = z_i)$, and take $\underline{t}^* = (t_1^*,\ldots,t_n^*)'$. Then equations (2.5) can be written in matrix form as

$$\partial L/\partial\underline{\tau} = \sum_{i=1} \underline{w}_i'(t_i^* - \underline{w}_i\underline{\tau}) = \underline{W}'\underline{t}^* - \underline{W}'\underline{W}\,\underline{\tau} = \underline{0}$$

and therefore as

$$\underline{W}'\underline{W}\,\underline{\tau} = \underline{W}'\underline{t}^* \quad . \tag{2.6}$$

If $\underline{t} = (t_1,\ldots,t_n)'$ were observable, it is well-known that the ML estimate of $\underline{\tau}$ would be the solution to the normal equations, $\underline{W}'\underline{W}\,\underline{\tau} = \underline{W}'\underline{t}$. If $\underline{\eta} = (\eta_1,\ldots,\eta_{m-1})'$ were known, the ML estimate of $\underline{\tau}$ would be the solution to (2.5), though since $\underline{t}^*$ as a function of $\underline{\tau}$, we would have to obtain this solution iteratively, e.g., by

$$\begin{aligned}\underline{\tau}^{(k+1)} &= (\underline{W}'\underline{W})^{-1}\underline{W}'\underline{t}^{*(k)} \\ &= \underline{\tau}^{(k)} + (\underline{W}'\underline{W})^{-1}\partial L/\partial\underline{\tau}\big|_{\underline{\omega} = [\underline{\eta}', \underline{\tau}^{(k)\prime}]'}\,,\end{aligned} \tag{2.7}$$

where $\underline{t}^{*(k)} = \underline{t}^*$ evaluated at $\underline{\tau} = \underline{\tau}^{(k)}$.

Equation (2.7) motivates the following choice for the matrix $\underline{G}^{(k)}$ in (2.4). We propose determination of $\underline{\nu}^{(k)}$ as a solution to the system

$$\begin{bmatrix} \underline{G}_1^{(k)} & \underline{G}_2^{(k)} \\ \underline{0} & \underline{W}'\underline{W} \end{bmatrix} \underline{\nu}^{(k)} = \partial L/\partial\underline{\omega}\big|_{\underline{\omega}=\underline{\omega}^{(k)}}\,, \tag{2.8}$$

where $[\underline{G}_1^{(k)}, \underline{G}_2^{(k)}]$ is taken to be either $-(\partial^2 L/\partial\underline{\eta}\partial\underline{\eta}', \partial^2 L/\partial\underline{\eta}\partial\underline{\tau}')$ or $-E(\partial^2 L/\partial\underline{\eta}\partial\underline{\eta}', \partial^2 L/\partial\underline{\eta}\partial\underline{\tau}')$, evaluated at $\underline{\omega} = \underline{\omega}^{(k)}$. The first (m-1) equations of (2.8) correspond to the first (m-1) equations of (2.4) for the NR or MS procedure, while the lower r equations of (2.8) come from (2.7). Computationally we proceed as follows:

1. Compute and store $(\underline{W}'\underline{W})^{-1}$.
2. At the k^{th} iteration, compute

$$\underline{\nu}_\tau^{(k)} = (\underline{W}'\underline{W})^{-1}\partial L/\partial\underline{\tau}\big|_{\underline{\omega}=\underline{\omega}^{(k)}}\,, \tag{2.9}$$

and then solve

$$\underset{\sim}{G}_1^{(k)} \underset{\sim}{\nu}_\eta^{(k)} = [\partial L/\partial \underset{\sim}{\eta}|_{\underset{\sim}{\omega}=\underset{\sim}{\omega}^{(k)}} - \underset{\sim}{G}_2^{(k)} \underset{\sim}{\nu}_\tau^{(k)}] \tag{2.10}$$

for $\underset{\sim}{\nu}_\eta^{(k)}$ to obtain $\underset{\sim}{\nu}^{(k)} = (\underset{\sim}{\nu}_\eta^{(k)'}, \underset{\sim}{\nu}_\tau^{(k)'})'$.

Having determined $\underset{\sim}{\nu}^{(k)}$, we determine $\underset{\sim}{\omega}^{(k+1)}$ using (2.3), choosing $\rho^{(k)}$ by the same approach used in conjunction with the NR and MS procedures.

We refer to this algorithm as the normal equation (NE) procedure. When $\Phi(\cdot)$ is the standardized normal CDF, the NE procedure is equivalent to the EM algorithm for determining $\underset{\sim}{\tau}$, together with a Newton-type procedure for determining $\underset{\sim}{\eta}$. (For discussion of EM algorithm see, e.g., Dempster, Laird, and Rubin, 1977.)

There is a potential advantage in determining $\underset{\sim}{\nu}^{(k)}$ by (2.8) relative to making the choices for $\underset{\sim}{\nu}^{(k)}$ associated with the NR or MS procedures. With (2.8) we do not have to solve an s dimensional system of equations during each iteration. Instead we compute $(\underset{\sim}{W}'\underset{\sim}{W})^{-1}$ initially and then, at each iteration, obtain $\underset{\sim}{\nu}_\tau^{(k)}$ from (2.9) and determine $\underset{\sim}{\nu}_\eta^{(k)}$ as a solution to an m-1 dimensional system of equations (2.10). Further, the coefficient matrix $\underset{\sim}{G}_1^{(k)}$ of (2.10) is tri-diagonal, so that the solution $\underset{\sim}{\nu}_\eta^{(k)}$ can be obtained quite simply (see, e.g., Westlake, 1975, pg. 34). Thus, when the number of parameters is very large and NR or MS is considered too costly, the NE procedure may provide an economical alternative.

Another motivation for the NE method is that $v\underset{\sim}{W}'\underset{\sim}{W}$ (for some constant v) approximates $-E[\partial^2 L/\partial\underset{\sim}{\tau}\partial\underset{\sim}{\tau}']$, since

$$-E[\partial^2 L/\partial\underset{\sim}{\tau}\partial\underset{\sim}{\tau}'] = \underset{\sim}{W}'\underset{\sim}{V}\underset{\sim}{W} ,$$

where $\underline{V}$ is a diagonal matrix with $(i,i)^{th}$ element $v_i = \sum_{j=1}^{m} (\phi_{i,j-1}-\phi_{ij})^2/P_{ij}$. When $\Phi(\cdot)$ is the standard normal CDF, it is readily verified, that $v_i = Var[E(t_i|Z_i)]$ and thus, using the relationship $Var(x) = Var[E(x|y)] + E[Var(x|y)]$,

$$v_i = Var(t_i) - E[Var(t_i|Z_i)] < 1.$$

$Var(t_i \mid Z_i = z_i)$ may be interpreted as the "loss of information" from observing Z_i rather than t_i. If $\underline{w}_i\underline{\tau}$ is inside the interval (η_1,η_{m-1}) and if $m > 3$, generally this loss of information will be small, and thus, $E[Var(t_i|Z_i)]$ will be near zero and v_i will be near the maximum vlaue of 1. Hence, we would have $\underline{W}'\underline{V}\underline{W} \doteq \underline{W}'\underline{W}$. Thus (2.9) may be viewed as an approximation to MS for determining $\underline{\nu}_{\tau}$. [If Φ is the logistic CDF, it can be shown (though the argument is nontrivial) that $v_i < (1/3)$. Hence

$$\underline{WVW} \doteq (1/3)\underline{W}'\underline{W} \tag{2.11}$$

when $m > 3$ and $\underline{w}_i\underline{\tau}$ is contained in (η_1,η_{m-1}) for most $i=1,\ldots,n$.]

4. EXAMPLES

We now present six examples. For each example, we compare the lack of fit when $\Phi(\cdot)$ is taken to be the standard normal or logistic CDF's. Table 5 displays this comparison for one model for each of the examples. We measure lack of fit by the likelihood ratio statistic (G^2), i.e., $-2\ln[\ell(\hat{\underline{\omega}})/\ell^*]$ where ℓ^* = the value of the likelihood function for a product multinomial distribution when the probabilities are taken to equal the observed frequencies, and by the Pearson chi-square statistic (X^2). For contingency tables, G^2 and X^2 are asymptotically distributed as χ^2 random variables (Bishop, Fienberg, and Holland, 1975, Section 4.9). Even for sparse data, the difference in the G^2 values for two hierarchical models may be well approximated by the χ^2 distribution if the two models differ by only a few parameters (Haberman, 1977). We use $-[E(\partial^2 L/\partial \underline{\omega} \partial \underline{\omega}')]^{-1}$ evaluated at $\hat{\underline{\omega}}$ to obtain estimated standard errors (se) for functions of $\hat{\underline{\omega}}$.

4.1 Preference for Black Olives

A survey was conducted among armed forces personnel concerning black olive preference. The respondents were classified by region of the U.S. and by urbanization (rural vs. urban). Preference was measured using a nine-category rating scale. We only have access to the "collapsed" data set (Table 1) having 6 ordered categories that was reported by Bock and Jones (1968).

The objective of this survey may have been to investigate the extent of differences in preference associated with region and urbanization (or perhaps to rank these subpopulations with respect to preference). We apply a threshold model in which the CDF is taken to be the logistic CDF and in which the linear model (2.1) is taken to be the 2-way classification (with interaction) model, where the two classifications correspond to urbanization (U) and region (R). Our parameterization is such that the first three columns of $\underline{W}$ are as in (2.2) while the elements of columns corresponding to the interaction effects are given by $w_{i4} = w_{i1}w_{i2}$ and $w_{i5} = w_{i1}w_{i3}$ for all i.

Model (Independent Variables)	r	G^2	df	P-value
1. ---	0	52.79	25	.00
2. U, R	3	24.15	22	.34
3. U, R, U x R	5	10.15	20	.51

Model 1 contains only the boundary points and corresponds to the restricted model with restrictions $\underline{\tau} = \underline{0}$. Model 2 is the 2-way classification (without interaction) model. The degrees of freedom (df) are computed as df = n*(m-1)-(m-1) - r, where n* = the number of "subpopulations". P-value denotes the probability of a χ_{df} random variable exceeding G^2. A small P-value indicates significant lack of fit.

The P-value for Model 1 indicates that the hypothesis $\underline{\tau} = \underline{0}$ is inconsistent with the data, i.e., there are differences among the subpopulations. Model 3 provides an adequate fit. The difference in G^2 between Models 2 and 3 is 5.00 (df = 5-3 = 2, P-value = .08). Thus there is some evidence that the U x R effect is needed. For Model 3, the subpopulations were ordered as follows:

Subpopulation	$\underline{w}_i\underline{\tau}$ ($\pm$ se)
rural/MW	$-.662 \pm .156$
rural/NE	$-.352 \pm .158$
urban/NE	$-.024 \pm .158$
urban/MW	$.209 \pm .155$
rural/SW	$.345 \pm .156$
urban/SW	$.485 \pm .153$

Those from the SW tend to like black olives more than those from the NE or MW. For each region (and particularly in the MW), preference tends to be higher for the urban group.

Bock applied a threshold model to these data, taking $\Phi(\cdot)$ to be the logistic CDF, but using a different parameterization for $\underline{W}\,\underline{\tau}$. He obtained estimates of the parameters using (noniterative) minimum χ^2 estimation. Bock reported a X^2 value for the 2-way classification (with interaction) model of 29.90 (P-value = .07), which is significantly poorer than that obtained using ML for the same model (X^2=18.35, P-value > .50).

Williams and Grizzle (1972) analyzed these data using what they called a scaling model. They state that their objective was to determine "whether subjects in different populations evaluate their preference for black olives according to the same scale (pg. 60)." Thus, they include parameters for "Population by (Preference) Category" interaction. Their model would coincide with Model 2 if the interaction with category was omitted. One criticism of their analysis is that, since they omit any U x R effects, U x R interaction being present could lead to the conclusion that the populations do not "evaluate their preference ... according to the same scale." (To include both U x R and all Population x Category effects would "saturate" their model, i.e., df = 0.) If the lack of fit for Model 3 were significant, instead of introducing Population x Category interaction into the model, we might consider some alternative to the logistic CDF.

Results obtained by taking Φ to be the standardized normal CDF were similar to those obtained for the logistic CDF (see Table 5). The differences between the estimates of the P_{ij}'s were all less than .007.

4.2 Extent of Current Drinking

Random samples of men were taken from the Bowery, a Manhattan skid row, from Camp LaGuardia, a rehabilitative institution near New York City, and from Park Slope, a lower-income Brooklyn neighborhood. Data were collected on the extent of recent drinking of alcoholic beverages and on the number of years of living in group quarters (institutional life) for each man interviewed. (Group quarters include armed forces barracks, prisons, hospitals, etc.) One hypothesis of the researcher (Bahr, 1969) was that extent of institutional life and current drinking are positively associated, e.g., that those who have had more than 5 years of institutional life would report greater consumption of alcohol. The data appear in Table 2.

We fit a threshold model to these data, taking current drinking as the response variable and regarding location (L) and number of years of institutional life (T) as classificatory variables. The threshold model fit slightly better when Φ was taken to be the standardized normal CDF than when it was taken to be the logistic CDF. The linear model for our analysis was the 2-way classification (without interaction) model, parameterized as described in Section 2. A summary of the fit obtained and the estimates is:

Model (Independent Variables)	r	G^2	df	P-Value
1. ---	0	56.58	16	.00
2. L	2	13.09	14	.52
3. L, T	4	8.17	12	.77

Parameter	Estimate
η_1	$-.317 \pm .061$
η_2	$.458 \pm .062$
τ_1	$.106 \pm .072$
τ_2	$.375 \pm .071$
τ_3	$-.116 \pm .069$
τ_4	$-.048 \pm .076$

[Bowery = τ_1; Camp LaGuardia = τ_2; Park Slope = $-\tau_1-\tau_2$; (0 years) = τ_3; (1-4 years) = τ_4; (5+ years) = $-\tau_3-\tau_4$].

As expected, the locations differ greatly with regard to the extent of current drinking, with Park Slope having a much lower average consumption. The estimates for τ_3 and τ_4 support the researcher's hypothesis, i.e., that the 5+ years group has a higher level of drinking $[(-\hat{\tau}_3-\hat{\tau}_4) - (\hat{\tau}_3+\hat{\tau}_4)/2 = .246 \pm .120]$.

Williams and Grizzle (1972) analyzed these data by assigning scores to (and averaging these scores over) the categories. They chose to assign a different set of scores for each location which, in effect, absorbed the large differences in extent of drinking among locations. Their analysis and the threshold model analysis coincided in the conclusion that the 5+ years group had a higher average consumption of alcohol. However, the threshold model analysis was more informative about differences among locations.

4.3 Pneumoconiosis Among Coal Miners

Survey data were obtained on the occurrence of the lung disease pneumoconiosis among coal miners. Each miner was classified into one of eight groups based on the number of years of mining and into one of three categories based on the degree of abnormality revealed by an X-ray. The data are given in Table 3. The number listed under "Years" is the midpoint of the group. Categories 1,2, and 3 represent normal X-ray, an initial stage of pneumoconiosis, and more severe stages of pneumoconiosis, respectively.

The logarithm (to the base 10) of years was taken to the independent variable. The lack of fit is not significant for either the standardized normal CDF or for the logistic CDF (but the G^2 value is 1.4 less for the normal CDF). The discrepancy between the two sets of estimated probabilities is greatest for the groups with the largest number of years. For 51.5 years the estimated probabilities of being in Category 1 under the normal and logistic models are .386 and .364, respectively (with standard errors $\doteq$ .054).

Andrich (1979) analyzed these data and proposed a model which he fit by pooling Categories 2 and 3. He concluded that once a decision on abnormality was made, the discrimination between Categories 2 and 3 was random and thus unrelated to years of mining. Yet considering the data in Categories 2 and 3 only, the observed relative frequencies for Category 3 are .33, .33, .62, .47, .53, .62, and .71. Except for the 27.5-year group, these are increasing (not random). Therefore, the implication of the threshold model that, when $\tau_1 > 0$, $\Pr(Z_i = 3|Z_i = 2 \text{ or } 3)$ is an increasing function of w_{i1} is consistent with these data. Andrich reports $X^2 = 8.15$ for his proposed model, whereas for the threshold model (which has one more parameter than Andrich's model) $X^2 = 3.52$.

4.4 Steer Nutrition

A study was conducted on the effects on yearling steers of the grain: silage ratio (R) in their feed and the intake level (I) of feed (restricted vs. unrestricted). Six silage: grain ratios were used, ranging from 93:7 to 0:100. Twelve pens (P) were used with 7-8 animals/pen. The experiment was replicated over 3 years (T). The coloring (C) of each animal was also recorded as belonging to one of 3 classes. To some extent, coloring reflects the breed of the animal. Two ordered categorical responses were the "quality" grade and "yield" grade of the carcass, with responses in 9 and 4 categories, respectively. Quality is a combination of marbling, texture, color, and age. Yield is the percentage of the steer usable for retail cuts. One objective was to determine the effect of level of intake on the grade of the carcass. The researchers were also interested in the effect on grade of R and in determining whether the effect for intake is constant over the levels of R.

We applied the threshold model to both the quality grade and yield grade data. With both data sets, we fit a 5-way classification model with first-order interactions involving T, I, and R. Due to the design of the experiment some of the main effects and interactions in the linear model were confounded with one another. Hence, the number of fixed effects for each model is less than if there had been no confounding. (The $\underline{W}$ matrix was obtained by deleting certain columns from what would be a full column rank model matrix were there no confounding.)

The following summarizes the fit for quality grade when Φ was taken to be the logistic CDF. Since, with sparse data, the χ^2 approximation for G^2 may be poor, we omit df's and P-values.

Model (Independent Variables)	r	G^2
1. ---	0	494.1
2. T,P,C,I,R	20	447.0
3. T,P,C,I,R,T*R	26	425.1
4. T,P,C,I,R,T*R,T*I,I*R	31	422.7

Model 3 is perhaps the best model, since the decrease in G^2 through the inclusion of the I*R and I*T effects is only 2.4. For Model 3, the estimate of the effect corresponding to I was 3.6 times its standard error (with unrestricted feed level being associated with higher quality). Since we observe no significant I*R interaction, we conclude that the increase in quality due to the change from restricted to unrestricted diet is roughly constant (on the underlying continuous scale) for different silage:grain ratios.

The magnitude of the τ_u's may be difficult to interpret. However, we can readily interpret the estimated probabilities from the threshold model analysis. Since price paid for a carcass is a function of quality grade (as well as yield grade, weight, etc.) we could, e.g., use the estimated probabilities (and the corresponding sale values) to estimate the expected increase in the dollar value of a carcass due to a change from restricted intake to unrestricted intake.

Quite different estimates for the probabilities were obtained for the quality grade data when Φ was taken to be the standardized normal CDF instead of the logistic CDF. Differences as great as .09 occurred in some of the estimated P_{ij}'s, e.g., for several observations, the estimated probability of falling in the middle category was .22 for the normal model vs. .31 for the logistic model.

4.5 Severity of Dumping Syndrome

Data were collected for 4 hospitals on the severity of dumping syndrome following surgery for ulcers. Severity was rated as none, slight or moderate. Four different surgical procedures were used. The objective of the study was to compare the effect of the surgical procedures. The data, as reported in Grizzle, Starmer, and Koch (1969), appear in Table 4.

We fit a two-way classification (without interaction) model to these data. A summary of the fit obtained is given in Table 5. We have included this example to provide an additional comparison of the iterative procedures.

5. COMPARISON OF ITERATIVE PROCEDURES

The MS, NR and NE procedures for obtaining estimates of the parameters of the normal threshold model were compared for each of the six data sets described in Section 4. The results are summarized in Table 6. In applying the NE procedure to the steer nutrition data (examples 4 and 5) we used $-[\partial^2 L/\partial\underline{\eta}\partial\underline{\omega}']$ for $[\underline{G}_1^{(k)}\underline{G}_2^{(k)}]$ in (2.8) since, with sparse data, observed second derivatives are computed more rapidly than expected second derivatives. In the other examples, we used $-E[\partial^2 L/\partial\underline{\eta}\partial\underline{\omega}']$ for $[\underline{G}_1^{(k)}\underline{G}_2^{(k)}]$.

The computations were performed by an ITEL AS/6 computer. We report the number of iterations (IT) until the % change in L became less than .005%. An asterisk indicates a case where a limit on the number of iterations was reached before the convergence criteria was satisfied. SEC denotes the amount of computer time in seconds. [SEC includes time for printing the first and second derivatives and the values of $\underline{\omega}$ and $L(\underline{\omega})$ at each iteration but does not include the time required to compute the estimated covariance matrix.] We also report the percentage by which the final L value differs from the maximum log-likelihood value (L_M) achieved using all three procedures.

The NR procedure consistently reached a value of L equal to or above the value obtained using the other procedures. For each example, the MS procedure reached essentially the same value for L(and $\hat{\underline{\omega}}$) as the NR procedure. However, MS required more time (and generally one more iteration) to converge than NR. For the quality-grade example, computing the ESD was especially time consuming since there were 9 categories and the data were sparse. For this example, MS took more than twice as much time as NR.

The NE procedure failed to converge within the specified number of iterations for examples 3 and 5. For two other examples, this procedure required more iterations than either MS or NR. For the quality grade example, the NE procedure did, in fact, require much less time per iteration than the MS or NR procedures. (With 28 parameters, the time required by MS and NR to determine $\underline{\nu}^{(k)}$ was quite large.) However, the NR procedure was quicker to converge.

The performance of the NR and MS procedures was virtually the same when Φ was taken to be the logistic CDF (as when Φ was the normal CDF). However, the NE procedure did not perform as well for example 6 when Φ was the logistic CDF. It required more than 40 iterations to converge (vs. 16 iterations when Φ was the normal CDF). Perhaps the NE procedure would perform better for the logistic distrituion if $\underline{\nu}_{\tau}^{(k)}$ were computed as

$$\underline{\nu}_{\tau}^{(k)} = 3(\underline{W}'\underline{W})^{-1} \left.\partial L/\partial \underline{\tau}\right|_{\underline{\omega}=\underline{\omega}^{(k)}}$$

[refer to (2.9) and (2.11)].

It was found that computing time could be reduced in most instances by only recomputing $\underline{G}^{(k)}$ in (2,4) or [$\underline{G}_1^{(k)}$ or $\underline{G}_2^{(k)}$] in (2.8) every i^{th} iteration, e.g., on the 1st, 4th, ... iterations. Although the number of iterations required for convergence tended to increase slightly, this increase tended to be outweighed by the savings from not having to recompute or "reinvert" the matrix of second derivatives on every iteration. For instance, if in the quality grade example the ESD are not recomputed after the first iteration, the MS procedure converges to the maximum in 7 iterations. The time required is only 2.85 seconds vs. 6.22 seconds when ESD was calculated at each iteration, even though the number of iterations is greater (7 vs. 4). For the NR procedure, convergence was reached in 1.60 seconds (vs. 2.76 seconds when SD was calculated at each iteration) and the number of iterations did not change. For the NE procedure, the time until termination (with a final L value .09% below L_M) was 3.01 seconds when $[\partial^2 L/\partial \underline{\eta} \partial \underline{\omega}']$ was recomputed only on the 1st and 10th iterations.

Table 1. Frequency of Preferences for Black Olives

Urbanization	Region	Category					
		1	2,3	4,5	6	7	8,9
Rural	NE	23	18	20	18	10	15
	MW	30	22	21	17	8	12
	SW	11	9	26	19	17	24
Urban	NE	18	17	18	18	6	25
	MW	20	15	12	17	16	28
	SW	12	9	23	21	19	30

Table 2. Extent of Current Drinking of Alcoholic Beverages

Location	Years of Institutional Life	Current Drinking: Light or Abstention	Moderate	Heavy
Bowery	0	25	21	26
	1-4	21	18	23
	5+	20	19	21
Camp LaGuardia	0	29	27	38
	1-4	16	13	24
	5+	8	11	30
Park Slope	0	44	19	9
	1-4	18	9	4
	5+	6	8	3

Table 3. Period of Exposure and Prevalence of Pneumoconiosis among a Sample of Coal Miners

Years	Degree of Abnormality		
	1	2	3
5.8	98	0	0
15.0	51	2	1
21.5	34	6	3
27.5	35	5	8
33.5	32	10	9
39.5	23	7	8
46.0	12	6	10
51.5	4	2	5

Table 4. Severity of Dumping Syndrome

Hospital	Surgical Procedure	Severity		
		None	Slight	Moderate
1	A	23	7	2
	B	23	10	5
	C	20	13	5
	D	24	10	6
2	A	18	6	1
	B	18	6	2
	C	13	13	2
	D	9	15	2
3	A	8	6	3
	B	12	4	4
	C	11	6	2
	D	7	7	4
4	A	12	9	1
	B	15	3	2
	C	14	8	3
	D	13	6	4

Table 5. Comparison of Normal and Logistic Threshold Models

Example	m	r	df	Normal			Logistic		
				L	G^2	X^2	L	G^2	X^2
1. Preferences for black olives	6	3	22	-1129.20	24.61	23.31	-1128.97	24.15	22.93
2. Extent of current drinking	3	4	12	-533.14	8.17	8.42	-533.97	8.67	8.91
3. Pneumoconiosis among coal miners	3	1	13	-203.57	3.61	3.52	-204.27	5.03	4.70
4. Steer nutrition: quality grade	9	31		-411.71	432.77		-406.65	422.65	
5. Steer nutrition: yield grade	4	31		-234.66	146.72		-234.64	146.68	
6. Dumping syndrome	3	6	24	-382.89	22.01	22.41	-382.76	21.74	22.07

Table 6. Comparison of Performance of Iterative Procedures

Example	m	r	Newton-Raphson IT	Newton-Raphson SEC	Newton-Raphson %below L_M	Method-of-scoring IT	Method-of-scoring SEC	Method-of-scoring %below L_M	Normal equation IT	Normal equation SEC	Normal equation %below L_M
1. Preference for black olives	6	3	2	.10	.00	3	.13	.00	3	.12	.00
2. Extent of current drinking	3	4	2	.09	.00	3	.12	.00	3	.13	.00
3. Pneumoconiosis among coal miners	3	1	5	.12	.00	5	.13	.00	25*	.42	8.66
4. Steer nutrition: quality grade	9	31	3	2.76	.00	4	6.22	.00	8	3.22	.11
5. Steer nutrition: yield grade	4	31	4	3.36	.00	4	4.38	.00	20*	5.58	.35
6. Dumping syndrome	3	6	2	.12	.00	3	.15	.00	16	.50	.00

6. REFERENCES

Aitchison, J. and Silvey, S.D. 1957. The generalization of probit analysis to the case of multiple responses. Biometrika 44, 131-140.

Anderson, J.A. 1980. Discussion of: Regression models for ordinal data. Journal of the Royal Statistical Society Ser. B 42, 130-131.

Andrich, D. 1979. A model for contingency tables having an ordered response. Biometrics 35, 403-415.

Ashford, J.R. 1959. An approach to the analysis of data for semi-quantal responses in biological assay. Biometrics 15, 573-581.

Bahr, H.M. 1969. Institutional Life, Drinking, and Disaffiliation. Social Problems 16, 365-375.

Bard, Y. 1974. Nonlinear Parameter Estimation. New York: Academic Press.

Berger, P.J. and Freeman, A.E. 1978. Prediction of sire merit for calving difficulty. Journal of Dairy Science 61, 1146-1150.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. 1975. Discrete Multivariate Analysis. Cambridge, Massachusetts: MIT Press.

Bock, R.D. 1975. Multivariate Statistical Methods in Behavioral Research. New York: McGraw Hill, Inc.

Bock, R.D. and Jones, L.V. 1968. The Measurement and Prediction of Judgement and Choice. San Francisco: Holden Day.

Brennan, R.W. 1978. Feedlot performance of carcass characteristics of yearling steers fed varying ratios of corn silage and corn grain. M.S. thesis. Iowa State University, Ames.

Dempster, A.P., Laird, N.M., and Rubin, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Ser. B 39, 1-38.

Grizzle, J.E., Starmer, C.F., and Koch, G.G. 1969. Analysis of categorical data by linear models. Biometrics 25, 489-504.

Guthrie, W.D., Russell, W.A., Reed, G.L., Hallauer, A.R., and Cox, D.F. 1978. Methods of evaluating maize for sheath-collar-feeding resistance to the European corn borer. Maydica 23, 45-53.

Haberman, S.J. 1977. Log-linear models and frequency tables with small expected cell counts. Annals of Statistics 5, 1148-1169.

Haberman, S.J. 1980. Discussion of: Regression models for ordinal data. Journal of the Royal Statistical Society Ser. B 42, 136-137.

Johnson, N.L. and Kotz, S. 1970. Continuous Univariate Distributions -1: Distributions in Statistics. New York: Wiley.

Koehler, K.J. and Larntz, K. 1980. An empirical investigation of goodness of fit statistics for sparse multinomials. Journal of the American Statistical Association 75, 336-344.

McCullagh, P. 1980. Regression models for ordinal data. Journal of the Royal Statistical Society Ser. B 42, 109-127.

McKelvey, R.D. and Zavoina, W. 1975. A statistical model for the analysis of ordinal level dependent variables. Journal of Mathematical Sociology 4, 103-120.

Mee, R.W. and Harville, D.A. 1981. Analysis of ordered categorial responses, assuming an underlying continuous variable. Submitted to the Journal of the American Statistical Association.

Snell, E.J. 1964. A scaling procedure for ordered categorical data. Biometrics 20, 592-607.

Westlake, J.R. 1975. A Handbook of Numerical Matrix Inversion and Solution of Linear Equations. Huntington, New York: Krieger.

Williams, O.D. and Grizzle, J.R. 1972. Analysis of contingency tables having ordered response categories. Journal of the American Statistical Association 67, 55-63.

PART III.

THRESHOLD MODEL WITH FIXED AND RANDOM EFFECTS

1. INTRODUCTION

The analysis of ordered categorical responses via the threshold approach is described in Mee and Harville (1981a,b). In those papers, the linear model for the assumed underlying continuous random variable is taken to be a fixed effects linear model. We now extend the threshold model approach by assuming that the underlying continuous random variable satisfies a mixed linear model, i.e., a linear model that includes random effects as well as fixed effects.

Procedures based on mixed linear models for a continuous response variable have been used extensively in animal breeding and plant breeding contexts. For example, in the analysis of milk production, it is common to assume a mixed linear model that includes a random sire effect. The purpose of such an analysis may be to 'predict' each sire effect and, perhaps, to estimate the variance of the sire effects. Mixed linear models have also been proposed for the analysis of calving difficulty (an ordered categorical response) for the purposes of evaluating sires regarding their merit with respect to calving ease and of estimating the sire variance to error variance ratio (e.g., Berger and Freeman, 1978).

We propose procedures for the analysis of ordered categorical responses via the threshold approach which are analogous to mixed linear model procedures for a continuous response variable. We discuss estimation of the fixed effects, prediction of the random effects, and estimation of variance ratios. Hence, by these procedures, the inferences which are generally of interest when one applies mixed linear model procedures to a continuous response variable may be made using a model for ordered categorical responses.

Other models which include random effects have been proposed for categorical responses. One approach is to express each response as a vector of indicator variables (with the j^{th} element equal to 1 if the response falls in the j^{th} category, and 0 otherwise) and then to apply some model for multivariate responses. Landis and Koch (1977), Quass and Van Vleck (1980), and Laird (1975) take this approach in the models they propose. These models do not utilize the category ordering and, therefore, seem less appropriate for settings where the categories do have a meaningful order. A second general approach is to assign numerical values to the categorical responses and then to analyze these scores via some linear model (see e.g., Berger and Freeman, 1978, and Tong, Wilton, and Schaeffer, 1977). Sometimes, with this approach, it is difficult to specify an appropriate linear model due to the violation (on the 'observed scale') of typical assumptions (Gianola, 1980). However, the primary objection to this approach seems to be that implicit in the assignment of numerical values is the inappropriate assumption that all responses within a category are identical.

In Section 2, the threshold model with random and fixed effects is presented. In Section 3, we indicate which functions of the parameters are identifiable and then, in Section 4.1, reparameterize the model in terms of a second threshold model [in a manner analogous to that described in Mee and Harville (1981a, Section 3.1) for the fixed-effects case]. In Section 4.2, we consider the form of the likelihood function for this alternative threshold model and note results which are useful for maximum likelihood (ML) estimation in this setting. Then, before proceeding further with our discussion of ML estimation for the threshold model, in Section 4.3 we review ML estimation procedures for the mixed linear model with a continuous response variable. Following this review, in Sections 4.4-4.6, we discuss the estimation of the parameters of the alternative threshold model and prediction of the random effects. The ML estimation procedures presented for the threshold model are extensions of ML estimation procedures for a continuous response variable. In Section 5, we illustrate the estimation and prediction procedures in terms of the analysis of calving-difficulty data and contrast these procedures with others which have appeared in the literature.

Subsequently, we refer to the model discussed by Mee and Harville (1981a,b) as the fixed effects threshold model.

2. MIXED EFFECTS THRESHOLD MODEL

Suppose the underlying continuous random response variable y_i for the i^{th} of n individuals or items satisfies the linear model

$$y_i = \underline{x}_i\underline{\alpha} + \underline{u}_i\underline{b} + e_i \qquad (i = 1,\ldots,n), \tag{3.1}$$

where

$\underline{x}_i = (x_{i1},\ldots,x_{ip})$ and $\underline{u}_i = (u_{i1},\ldots,u_{iq})$ are known vectors whose elements represent the values of the 'independent variables' associated with the i^{th} response,

$\underline{\alpha} = (\alpha_1,\ldots,\alpha_p)'$ is a vector of unknown parameters or fixed effects,

$\underline{b} = (b_1,\ldots,b_q)'$ is a vector of unknown random effects,

and

e_i = a random error.

We assume that the joint distribution of $(b_1,\ldots,b_q,e_1,\ldots,e_n)'$ is multivariate normal (MVN) with mean vector $\underline{0}$ and variance-covariance matrix $\sigma^2 \begin{bmatrix} \underline{D}(\underline{\rho}) & \underline{0} \\ \underline{0} & \underline{I} \end{bmatrix}$, where σ is strictly positive and the elements of the q x q matrix $\underline{D}(\underline{\rho})$ are assumed to be known functions of a (possibly unknown) vector of parameters $\underline{\rho} = (\rho_1,\ldots,\rho_c)'$.

Let $\underline{y} = (y_1,\ldots,y_n)'$, $\underline{X}$ = an n x p matrix with i^{th} row $\underline{x}_i$, and $\underline{U}$ = an n x q matrix with i^{th} row $\underline{u}_i$. We assume that rank($\underline{U}$) < n. The joint distribution of $\underline{y}$ is MVN with mean vector $\underline{0}$ and variance-covariance matrix $\sigma^2\underline{V}(\underline{\rho})$, where

$$\underline{V}(\underline{\rho}) = \underline{I} + \underline{U}[\underline{D}(\underline{\rho})]\underline{U}' \tag{3.2}$$

When the linear model (3.1) corresponds to an ordinary analysis of variance model, the vector $\underline{b}$ may be partitioned as

$$\underline{b} = (b_1',\ldots,b_c')' \tag{3.3}$$

in such a way that

$$\mathrm{Cov}(\underline{b}_i,\underline{b}_j) = \begin{bmatrix} \rho_i\underline{A}_i & i = j \\ \underline{0} & i \neq j \end{bmatrix},$$

where $\underline{b}_i$ is $q_i \times 1$ and $\underline{A}_i$ is a $q_i \times q_i$ known positive definite matrix. (Commonly the matrices $A_1,\ldots,A_c$ are taken to be identity matrices. However, we consider this more general case as occurs, e.g., in animal breeding applications when A_i is taken to be a matrix known as the relationship matrix.) Since $\rho_i\underline{A}_i$ is a variance-covariance matrix, we have that $\rho_i \geq 0$. Let $\underline{U} = (\underline{U}_1,\ldots,\underline{U}_c)$ be the partition of $\underline{U}$ correspoinding to (3.3). Then,

$$\underline{V}(\underline{\rho}) = \underline{I} + \sum_{i=1}^{c} \rho_i \underline{U}_i \underline{A}_i \underline{U}_i' \ . \tag{3.4}$$

When $\underline{V}(\underline{\rho})$ has the form (3.4), we refer to $\rho_1,\ldots,\rho_c$ as variance ratios.

The underlying continuous variable y_i cannot be observed. We can only observe that the i^{th} response falls into one of m categories. We number the categories $1,2,\ldots,m$ and take Z_i to be the discrete random variable defined by $Z_i = j$ when the i^{th} individual or item belongs to Category j. The relationship between y_i and Z_i is:

$$\xi_{j-1} < y_i \leq \xi_j \iff Z_i = j$$

where $j \in \{1,\ldots,m\}$, $\xi_o = -\infty$, $\xi_m = +\infty$, and $\underline{\xi} = (\xi_1,\ldots,\xi_{m-1})'$ is a vector of unknown boundaries. Let $\underline{\theta} = (\underline{\xi}', \underline{\tau}', \sigma, \underline{\rho}')'$, and take the parameter space for $\underline{\theta}$ to be

$$\textcircled{H} = \{\underline{\theta}: \xi_o < \xi_1 < \ldots < \xi_{m-1} < \xi_m,\ \underline{\tau} \in \underline{R}^r,\ \sigma > 0,\ \underline{\rho} \in P\}, \tag{3.5}$$

where P is a given subset of $\underline{R}^c$ such that, for all $\underline{\rho} \in P$, $\underline{D}(\underline{\rho})$ is positive definite or semidefinite.

Under these assumptions the joint probability mass function for $\underline{Z} = (Z_1,\ldots,\underline{Z}_n)'$ is

$$P[(z_1,\ldots,z_n)';\ \underline{\theta}] = \int\ldots\int_{\xi(\underline{z})} n[\underline{a};\ \underline{X\alpha},\ \sigma^2\underline{V}(\underline{\rho})]\ d\underline{a} \qquad (3.6)$$

$$(z_1=1,\ldots,m;\ldots;z_n=1,\ldots,m)$$

where

$$\xi(\underline{z}) = \{(a_1,\ldots,a_n)':\ \xi_{z_i-1} < a_i \leq \xi_{z_i},\ i = 1,\ldots,n\}$$

and

$n(\cdot;\underline{m},\underline{V})$ is the probability density function (PDF) for the MVN distribution with mean vector $\underline{m}$ and variance-covariance matrix $\underline{V}$.

We refer to the above model for $Z_1,\ldots,Z_n$ as the mixed effects threshold model.

In our analysis of calving difficulty, $\underline{b}$ will represent the sire effects and $\underline{D}(\underline{\rho})$ will have the form $\rho_1\underline{A}_1$. Using this model, we estimate ρ_1, which is of theoretical interest because it reflects the portion of total variability which is 'genetic' and determines the expected 'response due to selection' for calving ease. By our analysis, we obtain predictions for each of the random sire effects in order to identify the easy-calving sires. Calving difficulty is very evidently influenced by sex of calf, parity of the dam, and by herd-year-season. Through the inclusion of fixed effects for each of these factors, we hope to account for these differences and thus to effectively predict the sire effects.

3. IDENTIFIABILITY

The parameter vector $\underline{\theta}$ is not identifiable. However, we have the following lemma analogous to Lemma 1 of Mee and Harville (1981a). Define $\lambda_{ij}(\underline{\theta})$ by

$$\lambda_{ij}(\underline{\theta}) = (\xi_j - \underline{x}_i\underline{\alpha})/\sigma$$

and let $v_{ii'}(\underline{\rho})$ denote the $(i,i')^{th}$ element of $\underline{V}(\underline{\rho})$.

<u>Lemma 1</u>: When m>2, under the mixed effects threshold model, $\lambda_{ij}(\underline{\theta})$ $(i=1,\ldots,n;\ j=1,\ldots,m-1)$ and $v_{ii'}(\underline{\rho})$ $(i,i'=1,\ldots,n)$ are identifiable.

<u>Proof</u>: It suffices to show that for all $\underline{\theta}^+$ and $\underline{\theta}^* \in Ⓗ$, $P(\underline{z};\underline{\theta}^+) = P(\underline{z};\underline{\theta}^*)$ for all $\underline{z}$ implies $\lambda_{ij}(\underline{\theta}^+) = \lambda_{ij}(\underline{\theta}^*)$ and $v_{ii'}(\underline{\rho}^+) = v_{ii'}(\underline{\rho}^*)$, where $\underline{\theta}^+ = (\xi_1^+,\ldots,\xi_{m-1}^+, \underline{\tau}^{+'}, \sigma^+, \underline{\rho}^{+'})'$ and $\underline{\theta}^* = (\xi_1^*,\ldots,\xi_{m-1}^*, \underline{\tau}^{*'}, \sigma^*, \underline{\rho}^{*'})'$. Suppose that

$$P(\underline{z};\underline{\theta}^+) = P(\underline{z};\underline{\theta}^*) \quad \text{for all } \underline{z}. \tag{3.7}$$

Then, $\Pr(Z_i \leq j;\underline{\theta})$ is the same when $\underline{\theta} = \underline{\theta}^+$ as when $\underline{\theta} = \underline{\theta}^*$, implying that, for $i = 1,\ldots,n$,

$$\Phi[\lambda_{ij}(\underline{\theta}^+)/v_{ii}(\underline{\rho}^+)^{\frac{1}{2}}] = \Phi[\lambda_{ij}(\underline{\theta}^*)/v_{ii}(\underline{\rho}^*)^{\frac{1}{2}}] \ ,$$

where Φ is the standard normal CDF. Hence, we have that

$$\lambda_{ij}(\underline{\theta}^+)/v_{ii}(\underline{\rho}^+)^{\frac{1}{2}} = \lambda_{ij}(\underline{\theta}^*)/v_{ii}(\underline{\rho}^*)^{\frac{1}{2}} \quad (i=1,\ldots,n). \tag{3.8}$$

Further, using (3.8), we have that $\Pr(Z_i \leq j,\ Z_{i'} \leq j\ ;\underline{\theta})$ is the same when $\underline{\theta} = \underline{\theta}^+$ as when $\underline{\theta} = \underline{\theta}^*$, implying that, for $i \neq i' = 1,\ldots,n$,

$$F[t_{ij},\ t_{i'j}\ ;r_{ii'}(\underline{\rho}^+)] = F[t_{ij},\ t_{i'j}\ ;r_{ii'}(\underline{\rho}^*)] \ , \tag{3.9}$$

where $F(\cdot,\cdot;r)$ is the standardized bivariate normal CDF with correlation r, $r_{ii'}(\underline{\rho}) = v_{ii'}(\underline{\rho})/[v_{ii}(\underline{\rho})v_{i'i'}(\underline{\rho})]^{\frac{1}{2}}$, and $t_{ij} = \lambda_{ij}(\underline{\theta}^+)/v_{ii}(\underline{\rho}^+)^{\frac{1}{2}}$. Since, for fixed a and b, $F(a,b;r)$ is an increasing function of r (see, e.g., Owen, 1956, pg. 1078), (3.9) implies that

$$r_{ii'}(\underline{\rho}^{+}) = r_{ii'}(\underline{\rho}^{*}) \qquad (i \neq i' = 1,\ldots,n). \quad (3.10)$$

Equality (3.8) together with the identity

$$[\lambda_{ij}(\underline{\theta}) - \lambda_{ij'}(\underline{\theta})]/v_{ii}(\underline{\rho})^{\frac{1}{2}} = (\xi_j - \xi_{j'})/\sigma v_{ii}(\underline{\rho})^{\frac{1}{2}}$$

imply that

$$v_{ii}(\underline{\rho}^{+}) = c^2 v_{ii}(\underline{\rho}^{*}) \qquad (3.11)$$

where $c = (\sigma^*/\sigma^+)[(\xi_j^+ - \xi_{j'}^+)/(\xi_j^* - \xi_{j'}^*)]$ does not vary with i. Furthermore, (3.10) and (3.11) imply that

$$v_{ii'}(\underline{\rho}^{+}) = c^2 v_{ii'}(\underline{\rho}^{*}). \qquad (i,i' = 1,\ldots,n). \quad (3.12)$$

Hence, using expression (3.2) for $\underline{V}(\underline{\rho})$, $\underline{I} + \underline{UD}(\underline{\rho}^+)\underline{U}' = c^2[\underline{I} + \underline{UD}(\underline{\rho}^*)\underline{U}']$ and so $(c^2-1)\underline{I} = \underline{U}[\underline{D}(\underline{\rho}^+) - c^2\underline{D}(\rho^*)]\underline{U}'$. Finally,

$$\text{rank } [(c^2-1)\underline{I}] = \text{rank } \{\underline{U}[c^2\underline{D}(\underline{\rho}^*) - \underline{D}(\underline{\rho}^+)]\underline{U}'\} < n$$

[since rank$(\underline{U}) < n$], which implies $c^2 = 1$. Thus, by (3.12), $v_{ii'}(\underline{\rho}^+) = v_{ii'}(\underline{\rho}^*)$, and so by (3.8), $\lambda_{ij}(\underline{\theta}^+) = \lambda_{ij}(\underline{\theta}^*)$. []

Corollary 1: When $m>2$, if for $\underline{\rho}^+$ and $\underline{\rho}^* \epsilon P$, $\underline{UD}(\underline{\rho}^+)\underline{U}' = \underline{UD}(\underline{\rho}^*)\underline{U}'$ implies that $\underline{\rho}^+ = \underline{\rho}^*$, then the elements of $\underline{\rho}$ are identifiable.

By the change in variable $\underline{a}^* = \sigma^{-1}(\underline{a} - \underline{X\alpha})$, (3.6) may be expressed as

$$P(\underline{z};\underline{\theta}) = \int_{\lambda(\underline{z};\underline{\theta})} \cdots \int n[\underline{a}^*;\underline{0},\underline{V}(\underline{\rho})]\, d\underline{a}^* \qquad (3.13)$$

where

$$\lambda(\underline{z};\theta) = \{(a_1^*,\ldots,a_n^*)' : \lambda_{i,z_i-1}(\underline{\theta}) < a_i^* \leq \lambda_{iz_i}(\underline{\theta}),\ i=1,\ldots,n\}.$$

Hence $P(\underline{z};\underline{\theta})$ depends on $\underline{\theta}$ only through the values of $\lambda_{ij}(\underline{\theta})$ $(i=1,\ldots,n;$ $j=1,\ldots,m-1)$ and $\underline{V}(\underline{\rho})$. We have the following corollary to Lemma 1 analogous to Corollary 1 of Lemma 1 of Mee and Harville (1981a).

Corollary 2: Under the mixed effects threshold model, a parametric function is identifiable if and only if it can be expressed as a function of $\lambda_{ij}(\underline{\theta})$ $(i=1,\ldots,n;\ j=1,\ldots,m-1)$ and $\underline{V}(\underline{\rho})$.

In Mee and Harville (1981a), the probability mass function for $\underline{Z}$ under the fixed effects threshold model was re-expressed in terms of non-redundant identifiable parametric functions $\eta_1(\underline{\theta}),\ldots,\eta_{m-1}(\underline{\theta})$, $\tau_1(\underline{\theta}),\ldots,\tau_r(\underline{\theta})$. To obtain the analogous result for the mixed model, define $r = \text{rank}(\underline{1},\underline{X})$, let $\underline{W}$ be any $n \times r$ matrix such that the column spaces of $(\underline{1},\underline{W})$ and $(\underline{1},\underline{X})$ are identical and define $\eta_1(\underline{\theta}),\ldots,\eta_{m-1}(\underline{\theta})$, and $\underline{\tau}(\underline{\theta}) = [\tau_1(\underline{\theta}),\ldots,\tau_r(\underline{\theta})]'$ by

$$\underline{1}\eta_1(\underline{\theta}) - \underline{W}\underline{\tau}(\underline{\theta}) = \sigma^{-1}(\underline{1}\xi_1 - \underline{X}\underline{\alpha}) \qquad (3.14)$$

and

$$\eta_j(\underline{\theta}) - \eta_1(\underline{\theta}) = \sigma^{-1}(\xi_j-\xi_1) \quad (j=2,\ldots,m-1). \qquad (3.15)$$

Since the i^{th} element of the right hand side of (3.14) equals $\lambda_{i1}(\underline{\theta})$ and since $\sigma^{-1}(\xi_j-\xi_1) = \lambda_{1j}(\underline{\theta}) - \lambda_{11}(\underline{\theta})$, by Corollary 1, $\eta_1(\underline{\theta}),\ldots,\eta_{m-1}(\underline{\theta}),\tau_1(\underline{\theta}),\ldots,\tau_r(\underline{\theta})$ are identifiable functions of $\underline{\theta}$.

4. MAXIMUM LIKELIHOOD ESTIMATION

4.1 Alternative Mixed Effects Threshold Model

Consider an alternative mixed effects threshold model in which the underlying continuous variables $t_1,\ldots,t_n$ satisfy the linear model

$$t_i = \underline{w}_i\underline{\tau} + \underline{u}_i\underline{s} + d_i \qquad (i = 1,\ldots,n) \tag{3.16}$$

where

$\underline{w}_i$ = the i^{th} row of $\underline{W}$ (as defined in Section 3),

$\underline{\tau} = (\tau_1,\ldots,\tau_r)'$ is a vector of unknown parameters,

$\underline{s} = (s_1,\ldots,s_q)'$ is a vector of random effects,

and

$d_1,\ldots,d_n$ are independently distribution standard normal variates. We assume that $\underline{s}$ is distributed MVN with mean vector $\underline{0}$ and variance-covariance matrix $\underline{D}(\underline{\rho})$ and that $\mathrm{Cov}(\underline{s},d_i) = \underline{0}$ for all i. Hence, $\mathrm{Var}(t_1,\ldots,t_n) = \underline{V}(\underline{\rho})$ [as defined in 3.2)].

Under the alternative mixed effects threshold model, the relationship between t_i and the categorical response Z_i is assumed to be

$$\eta_{j-1} < t_i \leq \eta_j \iff Z_i = j,$$

where $j\varepsilon\{1,2,\ldots,m\}$, $\eta_0 = -\infty$, $\eta_m = +\infty$, and $\underline{\eta} = (\eta_1,\ldots,\eta_{m-1})'$ are the unknown boundaries for the alternative model. Let $\underline{\omega} = (\underline{\eta}',\underline{\tau}',\underline{\rho}')'$ and take the parameter space for $\underline{\omega}$ to be

$$\Omega = \{\underline{\omega}:\eta_0 < \eta_1 < \ldots < \eta_{m-1} < \eta_m,\ \underline{\tau}\ \varepsilon\ \underline{R}^r,\ \underline{\rho}\ \varepsilon\ P\}. \tag{3.17}$$

Under the alternative mixed effects threshold model, the joint probability mass function for $\underline{Z}$ is

$$P^*[(z_1,\ldots,z_n)';\underline{\omega}] = \int_{\eta(\underline{z})}\!\!\ldots\int n[\underline{a};\underline{W\tau},\ \underline{V}(\underline{\rho})]\ d\underline{a} \tag{3.18}$$

$$(z_1=1,\ldots,m;\ \ldots;\ z_n = 1,\ldots,m),$$

where

$$\eta(\underline{z}) = \{(a_1,\ldots,a_n)': \eta_{z_i-1} < a_i \le \eta_{z_i},\ i=1,\ldots,n\}. \tag{3.19}$$

Take $\underline{\omega}(\underline{\theta}) = [\eta_1(\underline{\theta}),\ldots,\eta_{m-1}(\underline{\theta}),\ \underline{\tau}'(\underline{\theta}),\ \underline{\rho}']'$. Since

$$\eta_j(\underline{\theta}) - \underline{w}_i\underline{\tau}(\underline{\theta}) = \sigma^{-1}(\xi_j - \underline{x}_i\underline{\alpha}) = \lambda_{ij}(\underline{\theta}),$$

we find, upon making the change of variables $\underline{a}^* = \underline{a} - \underline{W\tau}$ in the integral (3.18), that $P^*[\underline{z};\underline{\omega}(\underline{\theta})] = P(\underline{z};\underline{\theta})$ for all $\underline{z}$. Furthermore, the parameter spaces Θ and Ω are equivalent in the sense that the set $\{\underline{\omega}(\underline{\theta}): \underline{\theta} \in \Theta\}$ coincides with Ω. Hence, the elements of $\underline{\eta}$ and $\underline{\tau}$ are identifiable and, if the condition of Corollary 1 holds, then all of the parameters of the alternative mixed effects threshold model are identifiable.

We now discuss estimation of the parameter vector $\underline{\omega}$ of the alternative model. Due to the invariance property of ML, an ML estimate for $\underline{\omega}$ under the alternative mixed effects threshold model is also an ML estimate for the vector $\underline{\omega}(\underline{\theta})$ of functions of parameter vector $\underline{\theta}$ of the original model.

4.2. Likelihood Function and the EM Algorithm

The likelihood function (ℓ) for the alternative threshold model is

$$\ell(\underline{\omega}) = \int_{\eta(\underline{z})}\!\cdots\!\int n[\underline{a};\underline{W\tau},\ \underline{V}(\underline{\rho})]\ \underline{da}, \tag{3.20}$$

where $\eta(\cdot)$ is as defined in (3.19) and where $\underline{z}$ now denotes the vector of observed categorical responses. The ML estimate of $\underline{\omega}$ (if it exists) is defined to be an element $\hat{\underline{\omega}}$ of Ω such that

$$\ell(\hat{\underline{\omega}}) = \sup_{\underline{\omega}\varepsilon\Omega} \ell(\underline{\omega}) \ . \tag{3.21}$$

It will generally be infeasible to evalute (3.20) for a given value of $\underline{\omega}$ since the MVN integral cannot be reduced to the product of simple integrals. Thus we cannot use (3.20) itself as a basis for estimating $\underline{\omega}$. Before describing alternative procedures for estimating $\underline{\omega}$, we derive an alternative expression for $\ell(\underline{\omega})$ in order to exhibit a relationship between the fixed effects and mixed effects threshold models.

Conditioning on the random effects and using the relationship $f(t) = \int_{-\infty}^{\infty} g(t|s)h(s)ds$ [where $f(\cdot)$ and $h(\cdot)$ are marginal densities and $g(\cdot|\cdot)$ is a conditional density], we find that, when $\underline{D}(\underline{\rho})$ is nonsingular, $\ell(\underline{\omega})$ may be reexpressed as

$$\begin{aligned}\ell(\underline{\omega}) &= \int_{\eta(\underline{z})}\!\!\cdots\!\int \{\int_{-\infty}^{\infty}\!\!\cdots\!\int_{-\infty}^{\infty} n(\underline{a};\underline{W\tau} + \underline{Us}, \underline{I})n[\underline{s};\underline{0}, \underline{D}(\underline{\rho})]d\underline{s}\} \, d\underline{a} \\ &= \int_{-\infty}^{\infty}\!\!\cdots\!\int_{-\infty}^{\infty}\{\int_{\eta(\underline{z})}\!\!\cdots\!\int n(\underline{a};\underline{W\tau} + \underline{Us}, \underline{I})d\underline{a}\} \, n[\underline{s};\underline{0}, \underline{D}(\underline{\rho})] \, d\underline{s} \ .\end{aligned} \tag{3.22}$$

[The change in order of integration in (3.22) is permissible because the integrand is nonnegative.] The multiple integral inside the braces in (3.22) equals $\Pr(Z_1=z_1,\ldots,Z_n=z_n|\underline{s})$, which simplifies to

$$\prod_{i=1}^{n} \Pr(Z_i=z_i|\underline{s}) \ . \tag{3.23}$$

Note that when expression (3.23) is regarded as a function $\ell^*(\underline{\eta},\underline{\tau},\underline{s})$ of $\underline{\eta}$, $\underline{\tau}$, and $\underline{s}$, it corresponds to the likelihood function for a fixed effects model. We have that

$$\ell(\underline{\omega}) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \ell^*(\underline{\eta},\underline{\tau},\underline{s})n[\underline{s};\underline{0},\underline{D}(\underline{\rho})]\ d\underline{s}. \tag{3.24}$$

Hence, the likelihood function for the mixed effects threshold model is a weighted (by the PDF of $\underline{s}$) average of $\ell^*(\underline{\eta},\underline{\tau},\underline{s})$ values.

We now turn to the problem of estimating $\underline{\omega}$. In estimating $\underline{\tau}$ and $\underline{\rho}$, we make use of Dempster, Laird, and Rubin's (1977) results on the EM algorithm, which is an iterative algorithm for computing ML estimates which does not require that the likelihood function (or its derivatives) be evaluated. A brief summary of the EM algorithm and the relevent convergence properties follows.

To introduce the basic steps of the EM algorithm, suppose that $\underline{t} = (t_1,\ldots,t_n)'$ satisfying (3.16) is observed and suppose $\mathrm{Var}(\underline{s}) = \rho_1\underline{I}$. If the vector of random effects were observed, the ML estimate for ρ_1 would be $\underline{s}'\underline{s}/q$. Since $\underline{s}$ cannot be observed, we replace $\underline{s}'\underline{s}/q$ with its conditional expectation,

$$E[\underline{s}'\underline{s}/q\,|\,\underline{t}]\ , \tag{3.25}$$

as an estimate of ρ_1. However, (3.25) depends on the unknown ρ_1. Nevertheless, we might evaluate (3.25) at our best guess for ρ_1 in order to obtain a 'new' estimate.

The preceding discussion illustrates the intuitive idea behind the EM algorithm. We now give a more general description. Let $\underline{u}$ (unobservable) and $\underline{v}$(observable) be random vectors with densities $f(\underline{u};\pi)$ and $g(\underline{v};\pi)$ which depend on an unknown parameter π. Suppose

$$g(\underline{v};\pi) = \int_{R(\underline{v})} f(\underline{u};\pi)\, du ,$$

where $R(\underline{v})$ is a set (for $\underline{u}$) which depends on $\underline{v}$. Suppose that, given $\underline{u}$, $t(\underline{u})$ would be a sufficient statistic for π. Let $\pi^{(0)}$ be an initial value for π. When $f(\cdot;\pi)$ is from the regular exponential class of probability density functions, the value of π on the $(k+1)^{\underline{st}}$ cycle of the EM algorithm is the solution $\pi^{(k+1)}$ to

$$E_{\underline{u}}[t(\underline{u});\pi=\pi^{(k+1)}] = E_{\underline{u}|\underline{v}}[t(\underline{u})|\underline{v};\pi=\pi^{(k)}] .$$

Dempster, Laird, and Rubin (1977) showed that

$$L[\pi^{(k+1)}] \geq L[\pi^{(k)}] ,$$

where $L(\cdot)$ is the log-likelihood function for π, i.e., $L(\pi) = \ln[g(\underline{v};\pi)]$.

In each application of the EM algorithm, the observed response vector $\underline{v}$ is referred to as 'incomplete data', while the unseen random vector $\underline{u}$ which gives rise to the observed $\underline{v}$ [and from which the 'statistic' $t(\underline{u})$ is conceptualized] is referred to as 'complete data'. When we use the EM algorithm to estimate $\underline{\rho}$, $\underline{s}$ will be the complete data. When applying the EM algorithm to estimate $\underline{\tau}$ under the threshold model, the underlying continuous responses $t_1,\ldots,t_n$ will be the complete data.

Subsequently, we use the following notation. A superscript '(k)' will denote an estimate on the k^{th} iteration. Let x and y denote random variables, and let π denote a parameter. Expressions of the form

$E[y|x;\pi^{(k)}]$ and $Var[y|x;\pi^{(k)}]$ will be used to indicate the conditional expectation and conditional variance of y, given x, evaluated at the parameter value $\pi=\pi^{(k)}$.

4.3 ML Estimation when Continuous Responses are Observed

We review ML estimation for $\underline{\tau}$ and $\underline{\rho}$ when $t_1,\ldots,t_n$ satisfying (3.16) are observed. The discussion in this more familiar setting will assist our later discussion of ML estimation for the threshold model.

For the case where $\underline{D}(\underline{\rho})$ is known, the ML estimate for $\underline{\tau}$ is the solution $\underline{\tau}^o(\underline{\rho})$ to

$$\underline{W}'[\underline{V}(\underline{\rho})]^{-1}\underline{W}\,\underline{\tau}^o(\underline{\rho}) = \underline{W}'[\underline{V}(\underline{\rho})]^{-1}\underline{t} \tag{3.26}$$

(Searle, 1971, Section 10.8). (We use the argument $\underline{\rho}$ for $\underline{\tau}^o$ to emphasize that this solution depends on $\underline{\rho}$.) One difficulty in using (3.26) to determine $\underline{\tau}^o(\underline{\rho})$ is that the inverse of the n x n matrix $\underline{V}(\underline{\rho})$ must be computed. The solution $\underline{\tau}^o(\underline{\rho})$ may also be obtained using the equations [assuming that $\underline{D}(\underline{\rho})$ is nonsingular]

$$\begin{bmatrix} \underline{W}'\underline{W} & \underline{W}'\underline{U} \\ \underline{U}'\underline{W} & \underline{U}'\underline{U} + [\underline{D}(\underline{\rho})]^{-1} \end{bmatrix} \begin{bmatrix} \underline{\tau}^o(\underline{\rho}) \\ \underline{s}^o(\underline{\rho}) \end{bmatrix} = \begin{bmatrix} \underline{W}'\underline{t} \\ \underline{U}'\underline{t} \end{bmatrix} \tag{3.27}$$

(see e.g., Searle, 1971, Section 10.8). Equations (3.27), commonly known as 'Henderson's mixed model equations,' have the advantage over (3.26) of not requiring that $[\underline{V}(\underline{\rho})]^{-1}$ be computed. A second advantage is that $\underline{s}^o(\underline{\rho})$, the second part of the solution to (3.27), is related to

$$E[\underline{s}|\underline{t};\underline{\rho}] = \underline{D}(\underline{\rho})\ \underline{U}'[\underline{V}(\underline{\rho})]^{-1}(\underline{t} - \underline{W}\ \underline{\tau}). \tag{3.28}$$

In particular, $\underline{s}^{o}(\underline{\rho})$ equals (3.28) with $\underline{\tau}^{o}(\underline{\rho})$ substituted for $\underline{\tau}$ and is the 'best linear unbiased predictor' of $\underline{s}$ (see e.g., Henderson, 1975). In animal breeding contexts, the prediction $\underline{s}^{o}(\underline{\rho})$ is used as the estimated 'breeding value' for the animals represented in the vector $\underline{s}$.

The variance-covariance matrix of $\underline{\tau}^{o}(\underline{\rho})$ and $\underline{s}^{o}(\underline{\rho}) - \underline{s}$ (the 'error of prediction' for $\underline{s}$) is given by the inverse of the coefficient matrix of (3.27), i.e.,

$$\text{Var}\begin{bmatrix} \underline{\tau}^{o}(\underline{\rho}) \\ \\ \underline{s}^{o}(\underline{\rho})-\underline{s} \end{bmatrix} = \begin{bmatrix} \underline{W}'\underline{W} & \underline{W}'\underline{U} \\ \\ \underline{U}'\underline{W} & \underline{U}'\underline{U}+[\underline{D}(\underline{\rho})]^{-1} \end{bmatrix}^{-1} \tag{3.29}$$

[Refer to Henderson (1975) for the existence and properties of this inverse.]

We now consider the setting where $\underline{V}(\underline{\rho})$ has the form (3.4) and $\underline{\rho}$ (as well as $\underline{\tau}$) is a vector of unknown parameters. Let $\underline{s} = (\underline{s}_1', \ldots, \underline{s}_c')'$ denote the partitioning of $\underline{s}$ corresponding to (3.3), i.e., $\underline{s}_i$ is $q_i \times 1$ and $\text{Cov}(\underline{s}_i, \underline{s}_j) = \rho_i \underline{A}_i$ if $i = j$, and $\underline{0}$ otherwise. We chose to estimate $\underline{\rho}$ using the restricted maximum likelihood (REML) procedure (see e.g., Harville, 1977). We prefer REML to ML when r, the number of (unknown) fixed effects, is very large.

If the vector of random effects were observed

$$\underline{s}_i'\underline{A}_i^{-1}\underline{s}_i/q_i \tag{3.30}$$

would be the ML estimate of ρ_i. To obtain the REML estimate for ρ_i via the EM algorithm, on the $(k+1)^{st}$ iteration we compute the expectation of (3.30), conditional on $(\underline{I} - \underline{P}_w)\underline{t}$, where $\underline{P}_w = \underline{W}(\underline{W}'\underline{W})^{-1}\underline{W}'$, i.e.,

$$\rho_i^{(k+1)} = E\{\underline{s}_i'\underline{A}_i\underline{s}_i/q| \ (\underline{I} - \underline{P}_w)\underline{t} \ ; \ \underline{\rho}^{(k)}\} , \tag{3.31}$$

where $[\underline{\tau}^o(\underline{r})', \underline{s}^o(\underline{r})']'$ denotes the solution of (3.27) for $\underline{\rho} = \underline{r}$.

We now obtain an expression for (3.31) involving the conditional mean and conditional variance of $\underline{s}_i$. Let $\underline{x}$ be a random vector with mean $\underline{m}$ and variance-covariance matrix $\underline{V}$. Then, for any (conformable) constant matrix $\underline{A}$,

$$E(\underline{x}'\underline{Ax}) = \underline{m}'\underline{Am} + \text{trace } (\underline{A} \ \underline{V}) \ .$$

Now,

$$E[\underline{s}| \ (\underline{I} - \underline{P}_w)\underline{t} \ \ ;\underline{r}] = \underline{s}^o(\underline{r})$$

and

$$\text{Var}[\underline{s}| \ (\underline{I} - \underline{P}_w)\underline{t} \ \ ;\underline{r}] = \underline{C}(\underline{r}) \ ,$$

where $\underline{C}(\underline{r})$ = the lower right corner of (3.29) evaluated at $\underline{\rho} = \underline{r}$ (Harville, 1976). Using these results, we have that

$$\underline{\rho}_i^{(k+1)} = \{\underline{s}_i^{o'}[\underline{\rho}^{(k)}]\underline{A}_i^{-1}\underline{s}_i^o[\underline{\rho}^{(k)}] + \text{trace}[\underline{A}_i^{-1}\underline{C}_{ii}(\underline{\rho}^{(k)})]\}/q_i, \tag{3.32}$$

where $\underline{s}_i^o(\cdot)$ and $\underline{C}_{ii}(\cdot)$ denote the portions of $\underline{s}^o(\cdot)$ and $\underline{C}(\cdot)$ corresponding to $\underline{s}_i$. [For a detailed discussion of the EM algorithm applied to the estimation of $\underline{\rho}$, see Dempster, Rubin, and Tsutakwara (1981).] We note that the REML version of the EM algorithm is equivalent to the REML version of Henderson's algorithm for estimating variance components [compare our equation (3.32) with equation (6.3) of Harville (1977)].

To estimate $\underline{\tau}$ and $\underline{\rho}$ simultaneously, we adopt the iterative procedure whose $(k+1)^{st}$ iterate is obtained as follows:

1. Use $\underline{\rho}^{(k)}$ to compute $\underline{\tau}^{o}[\underline{\rho}^{(k)}]$ and $\underline{s}^{o}[\underline{\rho}^{(k)}]$ by equation (3.27).
2. Compute $\rho_i^{(k+1)}$ $(i=1,\ldots,c)$ using equation (3.32).
3. If, for some i, $|\rho_i^{(k+1)} - \rho_i^{(k)}|$ is greater than a constant δ, increase k by 1 and return to Step 1. Otherwise, stop.

This concludes our review for the case where $\underline{t}$ is observed.

4.4 ML Estimation for $\underline{\tau}$ and Prediction for $\underline{s}$

We describe a procedure for maximizing (3.20) with respect to $\underline{\tau}$ for the case where $\underline{\eta}$ and $\underline{\rho}$ are known quantities rather than unknown parameters. In doing so, we sometimes write $\underline{V}$ and $\underline{D}$ for $\underline{V}(\underline{\rho})$ and $\underline{D}(\underline{\rho})$, respectively. (Later we do describe estimation for $\underline{\tau}$, $\underline{\eta}$, and $\underline{\rho}$ jointly. However, for some applications, values for $\underline{\eta}$ and $\underline{\rho}$ available from some other source may be considered adequate for the purposes of the present analysis.)

If the vector $\underline{t}$ of underlying continuous responses were observed, then the ML estimate for $\underline{\tau}$ could be obtained using (3.26) or (3.27). When $\underline{t}$ is not observed, we apply the EM algorithm by computing the expected value of $\underline{t}$ (the complete data) conditional on $\underline{Z}$ (the incomplete data). Let

$$t_i^{(k)} = E[t_i|\underline{Z};\ \underline{\tau}^{(k)}] \qquad (i = 1,\ldots,n) \qquad (3.33)$$

and let $\underline{t}^{(k)} = [t_1^{(k)},\ldots,t_n^{(k)}]'$. We obtain the $(k+1)^{st}$ iterate of the EM algorithm for estimating $\underline{\tau}$ by simply replacing $\underline{t}$ with $\underline{t}^{(k)}$ in (3.27), i.e.,

$$\begin{bmatrix} \underline{W}'\underline{W} & \underline{W}'\underline{U} \\ \underline{U}'\underline{W} & \underline{U}'\underline{U} + \underline{D}^{-1} \end{bmatrix} \begin{bmatrix} \underline{\tau}^{(k+1)} \\ \underline{s}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \underline{W}'\underline{t}^{(k)} \\ \underline{U}'\underline{t}^{(k)} \end{bmatrix} . \tag{3.34}$$

In general, explicit expressions for the conditional expectations (3.33) required in (3.34) will not exist. We now obtain an alternative expression for (3.33) which suggests both an approximation for $\underline{t}_i^{(k)}$ and a predictor for the vector of random effects $\underline{s}$. Using the identity $E_x(x) = E_y[E_{x|y}(x|y)]$, where x and y denote random variables, we have that

$$E(t_i|\underline{Z}) = E_{\underline{s}|\underline{Z}}\{E[t_i|\underline{s},\underline{Z};\ \underline{\tau}]\} \quad . \tag{3.35}$$

Conditional on $\underline{s}$, t_i is normally distributed with mean $\underline{w}_i\underline{\tau} + \underline{u}_i\underline{s}$ and variance 1. Therefore,

$$E[t_i|\underline{s},\underline{Z};\ \underline{\tau}^{(k)}] = \underline{w}_i\underline{\tau} + \underline{u}_i\underline{s} + g_i(\underline{\tau},\underline{s}) \tag{3.36}$$

where

$$g_i(\underline{\tau},\underline{s}) = [\phi(\delta_{i,z_i-1}) - \phi(\delta_{i,z_i})]/[\Phi(\delta_{i,z_i}) - \Phi(\delta_{i,z_i-1})], \tag{3.37}$$

$\delta_{ij} = (\eta_j - \underline{w}_i\underline{\tau} - \underline{u}_i\underline{s})$ and $\phi(\cdot)$ is the standard normal PDF (Johnson and Kotz, 1970, pg. 81). Hence, from (3.33), (3.35) and (3.36),

$$t_i^{(k)} = \underline{w}_i\underline{\tau}^{(k)} + \underline{u}_i\hat{\underline{s}}[\underline{\tau}^{(k)}] + e_i[\underline{\tau}^{(k)}] \tag{3.38}$$

where

$$\hat{\underline{s}}(\underline{\tau}) = E[\underline{s}|\underline{Z};\underline{\tau}] \ . \tag{3.39}$$

and

$$e_i(\underline{\tau}) = E[g_i(\underline{\tau},\underline{s})|\underline{Z};\ \underline{\tau}] \ . \tag{3.40}$$

Using (3.38) to re-express $\underline{t}^{(k)}$, equations (3.34) may be rewritten as

$$\begin{bmatrix} \underline{W}'\underline{W} & \underline{W}'\underline{U} \\ \underline{U}'\underline{W} & \underline{U}'\underline{U}'+\underline{D}^{-1} \end{bmatrix} \begin{bmatrix} \underline{\tau}^{(k+1)} \\ \underline{s}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \underline{W}' \\ \underline{U}' \end{bmatrix} \{\underline{W\tau}^{(k)} + \underline{U\hat{s}}[\underline{\tau}^{(k)}] + \underline{e}[\underline{\tau}^{(k)}]\}, \quad (3.41)$$

where $\underline{e}(\underline{\tau}) = [e_1(\underline{\tau}),\ldots,e_n(\underline{\tau})]'$. Rearranging (3.41), we have that

$$\begin{bmatrix} \underline{W}'\underline{W} & \underline{W}'\underline{U} \\ \underline{U}'\underline{W} & \underline{U}'\underline{U}+\underline{D}^{-1} \end{bmatrix} \begin{bmatrix} \underline{\tau}^{(k+1)} - \underline{\tau}^{(k)} \\ \underline{s}^{(k+1)} - \underline{\hat{s}}^{(k)} \end{bmatrix} = \begin{bmatrix} \underline{W}'\underline{e}[\underline{\tau}^{(k)}] \\ \underline{U}'\underline{e}[\underline{\tau}^{(k)}]-\underline{D}^{-1}\underline{\hat{s}}[\underline{\tau}^{(k)}] \end{bmatrix}. \quad (3.42)$$

Thus, if $\underline{\tau}^{(k)}$ is such that the right hand side of (3.42) is null, then the solution to (3.41) has as its first component the ML estimate $\hat{\underline{\tau}}$ and as its second component $\underline{\hat{s}}(\hat{\underline{\tau}})$, the conditional mean (3.39) for $\underline{s}$ with $\hat{\underline{\tau}}$ substituted for $\underline{\tau}$. Hence, beginning with initial values $\underline{\tau}^{(0)}$ and $\underline{s}^{(0)}$ and computing, on the $(k+1)^{st}$ iterate, $\underline{\tau}^{(k+1)}$ and $\underline{s}^{(k+1)}$ as the solution to

$$\begin{bmatrix} \underline{W}'\underline{W} & \underline{W}'\underline{U} \\ \underline{U}'\underline{W} & \underline{U}'\underline{U}+\underline{D}^{-1} \end{bmatrix} \begin{bmatrix} \underline{\tau}^{(k+1)} - \underline{\tau}^{(k)} \\ \underline{s}^{(k+1)} - \underline{s}^{(k)} \end{bmatrix} = \begin{bmatrix} \underline{W}'\underline{e}[\underline{\tau}^{(k)}] \\ \underline{U}'\underline{e}[\underline{\tau}^{(k)}] - \underline{D}^{-1}\underline{s}^{(k)} \end{bmatrix}, \quad (3.43)$$

we may iteratively compute both $\hat{\underline{\tau}}$ and $\underline{\hat{s}}(\hat{\underline{\tau}})$.

We now address one final hurdle in the estimation of $\underline{\tau}$. The expected value (3.40) which is required in (3.43) has no explicit expression. To approximate $e_i(\underline{\tau})$, we initially propose using a first-order Taylor series approximation for $g_i(\underline{\tau},\underline{s})$. Expanding $g_i(\underline{\tau},\underline{s})$ about $\underline{\hat{s}}(\underline{\tau})$,

$$g_i(\underline{\tau},\underline{s}) \doteq g_i[\underline{\tau},\underline{\hat{s}}(\underline{\tau})] + [\partial g_i(\underline{\tau},\underline{s})/\partial \underline{s}]_{\underline{s}=\underline{\hat{s}}(\underline{\tau})}[\underline{s} - \underline{\hat{s}}(\underline{\tau})]. \qquad (3.44)$$

Taking the expectation of (3.44) with respect to the conditional contribution of $\underline{s}$, given $\underline{Z}$, we obtain $e_i(\underline{\tau}) \doteq g_i[\underline{\tau},\underline{\hat{s}}(\underline{\tau})]$, since the expected value of the second term is zero. Because we do not know $\underline{\hat{s}}(\tau)$, we suggest the further approximation

$$g_i[\underline{\tau}^{(k)},\underline{s}^{(k)}] \qquad (3.45)$$

for $e_i[\underline{\tau}^{(k)}]$.

Thus, we approximate the right hand side of (3.43) by

$$\begin{bmatrix} \underline{W}'g\ [\underline{\tau}^{(k)},\underline{s}^{(k)}] \\ \underline{U}'\underline{g}\ [\underline{\tau}^{(k)},\underline{s}^{(k)}] - \underline{D}^{-1}\underline{s}^{(k)} \end{bmatrix} \qquad (3.46)$$

where $\underline{g}[\underline{\tau},\underline{s}]$ denotes the n x 1 vector with i^{th} element $g_i(\underline{\tau},\underline{s})$.

Let $L^*(\underline{\eta},\underline{\tau},\underline{s}) = \ln[\ell^*(\underline{\eta},\underline{\tau},\underline{s})]$ [refer to the discussion following (3.23)]. We note that terms equal to $g_i(\underline{\tau},\underline{s})$ $(i=1,\ldots,n)$ [defined in (3.37)] appear in the first-order partial derivatives (with respect to the fixed effects) of the log-likelihood function for the fixed effects threshold model (refer to Mee and Harville, 1981b, Section 3.1). Expressing the first-order partial derivatives of $L^*(\underline{\eta},\underline{\tau},\underline{s})$ with respect to $\underline{\tau}$ and $\underline{s}$ in terms of $g_i(\underline{\tau},\underline{s})$, we have

$$\partial L^*(\underline{\eta},\underline{\tau},\underline{s})/\partial\tau_u = \sum_{i=1}^{n} w_{iu} g_i(\underline{\tau},\underline{s}) \quad (u=1,\ldots,r) \qquad (3.47)$$

and

$$\partial L^*(\underline{\eta},\underline{\tau},\underline{s})/\partial s_v = \sum_{i=1}^{n} u_{iv} g_i(\underline{\tau},\underline{s}) \quad (v=1,\ldots,q). \qquad (3.48)$$

Comparing (3.47) and (3.48) with the vector (3.46), we see that (3.46) equals

$$\begin{bmatrix} \partial L^*(\underline{\eta},\underline{\tau},\underline{s})/\partial\underline{\tau} \\ \partial L^*(\underline{\eta},\underline{\tau},\underline{s})/\partial\underline{s} - \underline{D}^{-1}\underline{s} \end{bmatrix} \tag{3.49}$$

evaluated at $\underline{\tau} = \underline{\tau}^{(k)}$ and $\underline{s} = \underline{s}^{(k)}$. [Note that this results does not require any conditions on the matrix $(\underline{W},\underline{U})$.]

Now denote the integrand of (3.24) (suppressing the dependence on $\underline{\rho}$) by $p(\underline{\eta},\underline{\tau},\underline{s})$. First we note that

$$\ln[p(\underline{\eta},\underline{\tau},\underline{s})] = L^*(\underline{\eta},\underline{\tau},\underline{s}) - \tfrac{1}{2}\,\underline{s}'\underline{D}^{-1}\underline{s} + K \tag{3.50}$$

(for some constant K). Hence, (3.49) is equivalent to the first-order partial derivatives of $\ln[p(\underline{\eta},\underline{\tau},\underline{s})]$ with respect to $\underline{\tau}$ and $\underline{s}$. Second, since $L^*(\underline{\eta},\underline{\tau},\underline{s})$ is a concave function of $\underline{\tau}$ and $\underline{s}$ (Burridge, 1981) and since $-\underline{s}'\underline{D}^{-1}\underline{s}$ is a concave function of $\underline{s}$, (3.50) is a concave function of $\underline{\tau}$ and $\underline{s}$. Hence, using (3.43) together with the approximation (3.45) for $e_i[\underline{\tau}^{(k)}]$ is equivalent to maximizing the integrand of $\ell(\underline{\omega})$ with respect to $\underline{\tau}$ and $\underline{s}$.

We propose replacing the coefficient matrix in (3.43) by the negative of the matrix of expected second-order partial derivatives of $\ln[p(\underline{\eta},\underline{\tau},\underline{s})]$ with respect to the elements of $\underline{\tau}$ and $\underline{s}$, evaluated at $\underline{\tau}^{(k)}$ and $\underline{s}^{(k)}$, i.e., on the $(k+1)^{st}$ iteration, determining new estimates $\underline{\tau}^{(k+1)}$ and $\underline{s}^{(k+1)}$ as the solution to

$$\begin{bmatrix} \underline{W}'\underline{R}^{(k)}\underline{W} & \underline{W}'\underline{R}^{(k)}\underline{U} \\ \underline{U}'\underline{R}^{(k)}\underline{W} & \underline{U}'\underline{R}^{(k)}\underline{U}+\underline{D}^{-1} \end{bmatrix} \begin{bmatrix} \underline{\tau}^{(k+1)}-\underline{\tau}^{(k)} \\ \underline{s}^{(k+1)}-\underline{s}^{(k)} \end{bmatrix} = \begin{bmatrix} \underline{W}'\underline{e}[\underline{\tau}^{(k)}] \\ \underline{U}'\underline{e}[\underline{\tau}^{(k)}]-\underline{D}^{-1}\underline{s}^{(k)} \end{bmatrix}, \tag{3.51}$$

where $\underline{R}^{(k)}$ is a diagonal matrix with $(i,i)^{th}$ element

$$r_i^{(k)} = \sum_{j=1}^{m} \{\phi[\delta_{i,j-1}^{(k)}] - \phi[\delta_{i,j}^{(k)}]\}^2 / \{\Phi[\delta_{i,j}^{(k)}] - \Phi[\delta_{i,j-1}^{(k)}]\}$$

and where $\delta_{i,j}^{(k)} = [\eta_j - \underline{w}_i \underline{\tau}^{(k)} - \underline{u}_i \underline{s}^{(k)}]$. The iterative procedure based on (3.51) will converge to the same point as would the procedure utilizing (3.43), since, with both systems, the coefficient matrix is positive definite and convergence is 'reached' when the right hand sides are null. We anticipate that the procedure based on (3.51) will generally converge in fewer iterations [than would the procedure based on (3.43)] since its coefficient matrix contains 'second derivative information.' When the approximation (3.45) is used, (3.51) corresponds to an iterate of the method-of-scoring procedure for maximizing $p(\underline{\eta},\underline{\tau},\underline{s})$ with respect to $\underline{\tau}$ and $\underline{s}$. Let

$$\underline{\Sigma}^{(k)} = \begin{bmatrix} \underline{W}'\underline{R}^{(k)}\underline{W} & \underline{W}'\underline{R}^{(k)}\underline{U} \\ \underline{U}'\underline{R}^{(k)}\underline{W} & \underline{U}'\underline{R}^{(k)}\underline{U} + \underline{D}^{-1} \end{bmatrix}^{-1} . \qquad (3.52)$$

When making inferences involving $\underline{\tau}$ and $\underline{s}$, we will take (3.52) from the final iteration as our estimate of the variance-covariance matrix of $\hat{\underline{\tau}}$ and $\hat{\underline{s}}(\hat{\underline{\tau}}) - \underline{s}$. [This matrix is the inverse of the information matrix of the 'posterior density,' $p(\underline{\eta},\underline{\tau},\underline{s})$.]

We also consider the use of a second approximation for $e_i(\underline{\tau})$. In place of (3.44) consider using a second-order Taylor series

$$g_i(\underline{\tau},\underline{s}) \doteq g_i[\underline{\tau},\hat{\underline{s}}(\underline{\tau})] + [\partial g_i(\underline{\tau},\underline{s})/\partial \underline{s}]_{\underline{s}=\hat{\underline{s}}(\underline{\tau})}[\underline{s} - \hat{\underline{s}}(\underline{\tau})]$$
$$+ \tfrac{1}{2}[\underline{s} - \hat{\underline{s}}(\underline{\tau})]'\underline{G}_i[\underline{s} - \hat{\underline{s}}(\underline{\tau})] , \qquad (3.53)$$

where $\underline{G}_i = [\partial^2 g_i(\underline{s},\underline{\tau})/\partial\underline{s}\partial\underline{s}']_{\underline{s}=\hat{\underline{s}}(\underline{\tau})}$, to approximate $g_i(\underline{\tau},\underline{s})$. Taking the expectation of (3.52) with respect to the conditional distribution of $\underline{s}$, given $\underline{Z}$, we obtain

$$e_i(\underline{\tau}) \doteq g_i[\underline{\tau},\hat{\underline{s}}(\underline{\tau})] + \tfrac{1}{2}\,\text{trace}\,\underline{G}_i\{\text{Var}[\underline{s} - \hat{\underline{s}}(\underline{\tau})|\underline{Z};\underline{\tau}]\}. \tag{3.54}$$

We propose using the lower right corner of (3.52) to approximate $\text{Var}[\underline{s} - \hat{\underline{s}}(\underline{\tau})|\underline{Z};\ \underline{\tau}]$ and to approximate $e_i[\underline{\tau}^{(k)}]$ by

$$g_i[\underline{\tau}^{(k)},\underline{s}^{(k)}] + \tfrac{1}{2}\,\text{trace}[\underline{G}_i^{(k)}\ \underline{\Sigma}_s^{(k)}], \tag{3.55}$$

where $\underline{G}_i^{(k)} = [\partial^2 g_i(\underline{s},\underline{\tau})/\partial\underline{s}\partial\underline{s}']$ evaluated at $\underline{\tau}=\underline{\tau}^{(k)}$ and $\underline{s}=\underline{s}^{(k)}$ and where $\underline{\Sigma}_s^{(k)}$ = the lower right corner of $\underline{\Sigma}^{(k)}$. Using $\underline{\Sigma}_s^{(k)}$ for $\text{Var}[\underline{s} - \hat{\underline{s}}(\underline{\tau})|\underline{Z};\underline{\tau}]$ in (3.54) is reasonable in that $\underline{\tau}_s^{(k)}$ is readily available and it accounts for the fact that $\underline{\tau}$ is being estimated. [In the continuous case, to account for the fact that $\underline{\tau}$ is estimated, the lower-right corner of (3.29) is generally used as the conditional variance of $\underline{s}$, given $\underline{t}$ (see, e.g., Harville, 1976, and Dempster, Rubin, and Tsutakawa, 1981).]

In the Appendix, we compare the first and second approximations for $e_i[\tau^{(k)}]$, i.e. (3.45) and (3.55), with $e_i[\underline{\tau}^{(k)}]$ in settings where $e_i[\underline{\tau}^{(k)}]$ can be computed. We also evaluate the accuracy of our estimate for the conditional variance of $\underline{s}$. Further comparisons are provided in Section 5.

We make one final observation regarding the estimation of the fixed effects. Suppose some of the elements of $\underline{\tau}$ represent 'block' effects. If all the responses in a block fall in Category 1 or all in Category m, then $\ell^*(\underline{\eta},\underline{\tau},s)$ is a monotone function of that block effect (Mee and Harville, 1981a). Hence, $\ell(\underline{\omega})$ is also a monotone function of that block effect and the ML estimate for that effect will not exist. Hence, before computing ML estimates for the remaining parameters, we delete all observations which appear in blocks for which all the responses are in Category 1 or all in Category m. The justification for this procedure is the same as with the fixed effect threshold model (refer to Mee and Harville, 1981a).

4.5 Estimation of η

Suppose now that $\underline{\rho}$ is a known quantity, but that $\underline{\eta}$ and $\underline{\tau}$ are unknown parameters to be estimated. In the previous section, we described procedures based on the EM algorithm for computing the ML estimate of $\underline{\tau}$ without ever evaluating the likelihood function (3.20). For the estimation of $\underline{\eta}$, we see no way to apply the EM algorithm. [Aitkin has agreed with the conclusion that the EM algorithm cannot be applied to the estimation of the boundaries for even the fixed effects threshold model (Aitkin, 1981).]

We propose as an estimate for $\underline{\eta}$ the vector $\hat{\underline{\eta}}$ which will maximize $p[\underline{\eta},\hat{\underline{\tau}},\hat{s}(\hat{\underline{\tau}})]$. (The estimate $\hat{\underline{\eta}}$ is not the ML estimate of $\underline{\eta}$ under the mixed effects threshold model, but it would be the ML estimate if the underlying model were a fixed effects model.) Since $\hat{\underline{\tau}}$ and $\hat{\underline{s}}(\hat{\underline{\tau}})$ depend on $\underline{\eta}$, $\hat{\underline{\eta}}$, $\hat{\underline{\tau}}$, and $\hat{\underline{s}}(\hat{\underline{\tau}})$ must be computed 'simultaneously'.

When the first approximation for $e_i(\underline{\tau})$ is used,

$$p[\hat{\underline{\eta}},\hat{\underline{\tau}},\hat{\underline{s}}(\hat{\underline{\tau}})] = \max_{\underline{\eta},\underline{\tau},\underline{s}} p(\underline{\eta},\underline{\tau},s) \tag{3.56}$$

[refer to discussion following (3.50)]. When the second approximation for $e_i(\underline{\tau})$ is used to obtain $\hat{\underline{\tau}}$ and $\hat{\underline{s}}(\hat{\underline{\tau}})$, (3.56) will not necessarily hold.

To estimate $\underline{\eta}$ and $\underline{\tau}$, we propose the iterative procedure for which, on the $(k+1)^{st}$ iteration, we compute $\underline{\tau}^{(k+1)}$ and $\underline{s}^{(k+1)}$ by (3.51) and then compute $\underline{\eta}^{(k+1)}$ as the solution to

$$-[\underline{G}_{\eta\eta}^{(k)} \; \underline{G}_{\eta\tau}^{(k)} \; \underline{G}_{\eta s}^{(k)}] \begin{bmatrix} \underline{\eta}^{(k+1)}-\underline{\eta}^{(k)} \\ \underline{\tau}^{(k+1)}-\underline{\tau}^{(k)} \\ \underline{s}^{(k+1)}-\underline{s}^{(k)} \end{bmatrix} = \partial L^*(\underline{\eta},\underline{\tau},\underline{s})/\partial\underline{\eta}\Big|_{(\underline{\eta},\underline{\tau},s)=[\underline{\eta}^{(k)},\underline{\tau}^{(k)},\underline{s}^{(k)}]} \tag{3.57}$$

where

$$\underline{G}_{\eta\eta}^{(k)} = \partial^2 L^*(\underline{\eta},\underline{\tau},\underline{s})/\partial\underline{\eta}\partial\underline{\eta}'\Big|_{(\underline{\eta},\underline{\tau},\underline{s})=[\underline{\eta}^{(k)},\underline{\tau}^{(k)},\underline{s}^{(k)}]}$$

and where $\underline{G}_{\eta\eta}^{(k)}$ and $\underline{G}_{\eta s}^{(k)}$ are similarly defined.

4.6 ML Estimation for Variance Ratios

Suppose that $\underline{V}(\underline{\rho})$ is as given in (3.4) and that $\underline{\eta}$, $\underline{\tau}$ and $\underline{\rho}$ are to be estimated. (Since ρ_i represented a variance ratio under the original mixed effects model, we continue to refer to it in this way.) The REML version of the EM algorithm for estimating ρ_i when $\underline{t}$ is observed was described in Section 4.3 [see (3.32)]. When the continuous responses are not observed, $\underline{Z}$ instead of $\underline{t}$ (or, in the REML version, $(\underline{I} - \underline{P}_W)\underline{t}$) takes the role of the incomplete data. Thus, we must compute the conditional expectation of (3.30), given $\underline{Z}$, i.e., on the $(k+1)^{st}$ iterate,

$$\rho_i^{(k+1)} = E[\underline{s}_i'\underline{A}_i^{-1}\underline{s}_i|\underline{Z};\underline{\rho}^{(k)}]/q_i \tag{3.58}$$

$$= \{E[\underline{s}_i'|\underline{Z};\ \underline{\rho}^{(k)}]\underline{A}_i^{-1}E[\underline{s}_i|\underline{Z};\underline{\rho}^{(k)}] + \mathrm{trace}[\underline{A}_i^{-1}\mathrm{Var}(\underline{s}_i|\underline{Z};\underline{\rho}^{(k)})]\}/q_i$$

We will make use of the following notation. Let $\hat{\underline{\tau}}^{(k)}$, $\hat{\underline{s}}^{(k)}$ and $\hat{\underline{\eta}}^{(k)}$ denote $\hat{\underline{\tau}}$, $\hat{\underline{s}}(\hat{\underline{\tau}})$, and $\hat{\underline{\eta}}$ computed at $\underline{\rho}=\underline{\rho}^{(k)}$ (using the procedures of Sections 4.4 and 4.5). Also, let $\hat{\underline{\Sigma}}^{(k)}$ denote the inverse of the negative of the matrix of expected second-order partial derivatives of $\ln[p(\underline{\eta},\underline{\tau},\underline{s})]$ with respect to $\underline{\tau}$ and $\underline{s}$, evaluated at $[\hat{\underline{\eta}}^{(k)},\hat{\underline{\tau}}^{(k)},\hat{\underline{s}}^{(k)}]$, i.e.,

$$\begin{bmatrix} \underline{W}'\hat{\underline{R}}^{(k)}\underline{W} & \underline{W}'\hat{\underline{R}}^{(k)}\underline{U} \\ \underline{U}'\hat{\underline{R}}^{(k)}\underline{W} & \underline{U}'\hat{\underline{R}}^{(k)}\underline{U} + \{\underline{D}[\hat{\underline{\rho}}^{(k)}]\}^{-1} \end{bmatrix}^{-1} \tag{3.59}$$

where $\underline{R}^{(k)}$ is a diagonal matrix with $(i,i)^{th}$ element

$$\hat{r}_i^{(k)} = \sum_{j=1}^{m} \{\phi[\hat{\delta}_{i,j-1}^{(k)}] - \phi[\hat{\underline{\delta}}_{i,j}^{(k)}]\}^2/\{\Phi[\hat{\delta}_{i,j}^{(k)}] - \Phi[\hat{\delta}_{i,j-1}^{(k)}]\}$$

and where $\hat{\delta}_{i,j}^{(k)} = [\hat{\eta}_j^{(k)} - \underline{w}_i\hat{\underline{\tau}}^{(k)} - \underline{u}_i\hat{\underline{s}}^{(k)}]$.

On the $(k+1)^{st}$ iterate, we compute

$$\rho_i^{(k+1)} = \{\hat{\underline{s}}_i^{(k)\,'}\underline{A}_i^{-1}\hat{\underline{s}}_i^{(k)} + \mathrm{trace}[\underline{A}_i^{-1}\hat{\underline{\Sigma}}_i^{(k)}]\}/q_i, \tag{3.60}$$

where $\hat{\Sigma}_i^{(k)}$ = the $q_i \times q_i$ portion corresponding to $\underline{s}_i$ of (3.59). Equation (3.60) is analogous to (3.32), the REML version of the EM algorithm for ρ_i in the continuous case in that our estimate $\hat{\underline{\Sigma}}_i^{(k)}$ for the conditional variance of $\underline{s}_i$ accounts for the fact that $\underline{\tau}$ is unknown and must be estimated.

To estimate $\underline{\eta}$, $\underline{\tau}$, and $\underline{\rho}$ simultaneously, we adopt the iterative procedure whose $(k+1)^{st}$ iterate is obtained as follows:

1'. Use $\underline{\rho}^{(k)}$ to compute $\hat{\underline{\eta}}^{(k)}$, $\hat{\underline{\tau}}^{(k)}$, and $\hat{\underline{s}}^{(k)}$.

2'. Compute $\rho_i^{(k+1)}$ $(i=1,\ldots,c)$ using (3.60).

3'. If, for some i, $|\rho_i^{(k+1)} - \rho_i^{(k)}| > \delta$, where δ is a chosen constant, increase k by 1 and return to Step 1'. Otherwise, stop.

Steps 2' and 3' above are essentially the same as Steps 2 and 3 of the iterative procedure described in Section 4.3 for the continuous case. However, Step 1' is more difficult than Step 1 of Section 4.3 in two respects. First, Step 1' involves estimation of $\hat{\underline{\eta}}^{(k)}$ using the system (3.57). Second, for each k, Step 1' itself requires an iterative procedure, i.e., $\hat{\underline{\eta}}^{(k)}$, $\hat{\underline{\tau}}^{(k)}$, and $\hat{\underline{s}}^{(k)}$ must be computed iteratively using systems (3.51) and (3.57), whereas, for Step 1, $\underline{\tau}^o[\underline{\rho}^{(k)}]$ and $\underline{s}^o[\underline{\rho}^{(k)}]$ may be computed directly using (3.27).

5. EXAMPLE: CALVING DIFFICULTY

We now apply the procedures of Section 4 to the analysis of calving-difficulty data. Our objectives in applying the threshold model to these data are: 1) estimation of the sire variance to error variance ratio, 2) prediction of the sire effects and 3) estimation, for each sire, of the probability of difficult births from heifers. We analyze two data sets. In Section 5.1, we analyze the data which appeared in Quass and Van Vleck (1980) and contrast the threshold approach with the approach discussed in that paper. In Section 5.2, we analyze data that were collected by American Breeders Service (ABS). In addition to the threshold model analysis, we analyze the ABS data using the approach described by Berger and Freeman (1978).

With each application of the threshold model, we compute estimates for the parameters using both the first approximation (1st A) and second approximation (2nd A), i.e., (3.45) and (3.55), for $e_i[\underline{\tau}^{(k)}]$.

5.1 Quass and Van Vleck

The data analyzed by Quass and Van Vleck (1980) appear in Table 1. These data represent calving difficulty experienced by Holstein heifers giving birth for the first time with the calf being a female. To simplify the analysis, they ignored HYS effects. Categories 1,2, and 3 denote no difficulty, slight difficulty, and extreme difficulty, respectively.

Table 1. Calving Difficulty by Sire

	Category		
Sire	1	2	3
1	50	30	20
2	40	10	0
3	140	50	10
4	130	15	5

We analyzed these data using the procedures described in Section 4. The underlying model (3.1) was taken to be the random effects model,

$$y_i = \alpha_1 + \underline{u}_i\underline{b} + e_i \qquad (i=1,\ldots,n),$$

where $\underline{u}_i$ = a 4 x 1 vector of indicator variables. Hence, the underlying model (3.16) for the alternative threshold model is

$$t_i = \underline{u}_i\underline{s} + d_i \qquad (i=1,\ldots n).$$

We assume that the sires are unrelated and, thus, take $\mathrm{Var}(\underline{s}) = \rho_1\underline{I}$, with ρ_1 unknown.

Using the procedures of Section 4, we obtained:

Parameter	Estimate	
	1st A	2nd A
η_1	.6125	.5850
η_2	1.5778	1.5506
ρ_1	.1702	.1723

The predictions of the sire effects for each approximation were:

Sire	Prediction	
	1st A	2nd A
1	.6105	.5851
2	-.2409	-.2698
3	.0566	.0293
4	-.4262	-.4556

The same ranking of the sires is provided by the two approximations. The differences between the predictions obtained using the two approximations may be accounted for by the differences in the estimates of the boundaries. In Table 2, the predicted frequencies for each sire are given. The differences are all less than .001.

Table 2. Predicted Frequencies

Sire	1st A Categories 1	2	3	2nd A Categories 1	2	3
1	.5008	.3325	.1667	.5000	.3329	.1671
2	.8033	.1622	.0345	.8037	.1620	.0343
3	.7109	.2250	.0641	.7108	.2251	.0641
4	.8505	.1269	.0225	.8510	.1266	.0224

To verify the accuracy of the 2nd A for the purposes of estimating ρ_1 and $E(\underline{s}|\underline{Z};\rho_1)$, we computed $E(\underline{s}|\underline{Z};\rho_1)$ and $Var(\underline{s}|\underline{Z};\rho_1)$ numerically, taking ρ_1, η_1, and η_2 equal to the estimated values obtained using the 2nd A. In Table 3, we contrast the approximations for $E(s_i|\underline{Z};\hat{\rho}_1)$ and $[Var(s_i|\underline{Z};\hat{\rho}_1)]^{\frac{1}{2}}$ with the values obtained numerically. [The approximation for the conditional variance of s_i, given $\underline{Z}$, was obtained using (3.52).]

Table 3. Approximate vs. Exact Predictions and Standard Deviations

i	$E(s_i\|\underline{Z};\hat{\rho}_1)$ 2nd A	Exact	$[Var(s_i\|\underline{Z};\hat{\rho}_1)]^{\frac{1}{2}}$ 2nd A	Exact
1	.5851	.5851	.1106	.1108
2	−.2698	−.2699	.1782	.1780
3	.0293	.0293	.0879	.0876
4	−.4556	−.4557	.1174	.1190

The agreement between exact and approximate values in Table 3 is excellent. Using the exact $E(s_i|\underline{Z};\hat{\rho}_1)$ and $Var(s_i|\underline{Z};\hat{\rho}_1)$ values to 're-estimate' ρ_1, we obtain [using (3.60)] $\hat{\rho}_1$ = .1724 (which differs from the estimate obtained using the 2nd A by only .0001).

As noted in the introduction, the approach taken by Quass and Van Vleck does not assume that the categories are ordered. With their approach, a multivariate analysis of variance model is assumed for the m-1 vector of indicator variables (with the j^{th} element equal to 1 if the response is in Category j; j=1,...,m-1). Estimated frequencies for each category are then computed simply as a linear combination of the estimated effects. (There is a possibility of some estimated frequencies falling outside of the interval [0,1] unless constraints are incorporated in the estimation procedure.) Each sire is represented by an m-1 vector of random effects. Hence, when $m > 2$, this approach does not necessarily provide a unique ranking of the sires. This multivariate model is even less suited for estimating the ratio of the sire variance to error variance, since these variances are (m-1) x (m-1) matrices.

5.2 American Breeders Service

Data on calving difficulty were collected by American Breeders Service from herds using semen from ABS Holstein sires. These data, collected from 1976-1980, were made available through the Department of Animal Science, Iowa State University.

The U.S.D.A. Active Sire Summary List for Summer 1980 (North Central Region Extension Publication) listed 108 active Holstein sires for ABS. Eighty-five of these appeared in data available to us. From the ABS data, we selected for our analysis only the records for these 85 active sires. There were 13,352 such records.

Calving difficulty was recorded using a five category scale. The categories and the observed frequencies were:

Category	Frequency	Percentage
1. No problem	10,516	78.8
2. Slight difficulty	949	7.1
3. Needed assistance	1,170	8.8
4. Considerable force needed	482	3.6
5. Extreme difficulty	235	1.8

In addition to calving difficulty and sire of calf, parity of the dam, sex of calf, and HYS were recorded. Also, the sire and maternal grandsire for each of the 85 bulls having progeny data were known.

We analyzed these data using the threshold approach. The assumed linear model for calving difficulty on the underlying scale included fixed effects for HYS, sex, and parity, and a random effect for sire. The variance-covariance matrix for the sire effects was taken to be $\rho_1\underline{A}_1$, where $\underline{A}_1$ = the relationship matrix (see, e.g., Falconer, 1960, pg. 233) for the 85 bulls. [The relationship matrix was computed assuming that the sires and maternal grandsires (of the bulls with progeny data) not included among the 85 active sires were unrelated.]

The 13,352 records appeared in 1967 HYS's. There were 866 HYS's for which all the responses were in Category 1 or all were in Category 5. The 2975 observations in these 866 HYS were deleted (refer to the discussion at the close of Section 4.4) and the remaining 10,377 records (1101 HYS's) were used in the analysis.

To satisfy the requirement that $(\underline{1},\underline{W})$ have full column rank, we parameterized as follows. Define a 10,377 x 1101 matrix $\underline{H}$ with $(i,j)^{th}$ element h_{ij} defined as

$$h_{ij} = \begin{cases} 1, & \text{if the } i^{th} \text{ record is in the } j^{th} \text{ HYS} \\ 0, & \text{otherwise} \end{cases} .$$

The first column of $\underline{H}$ was deleted and the remaining matrix (with 1100 columns), plus 3 additional columns for the other fixed effects, composed the matrix $\underline{W}$. The elements of the other 3 columns were given by

$$(w_{i,1101}, w_{i,1102}) = \begin{cases} (1,0), & \text{if parity} = 1 \text{ for the } i^{th} \text{ record} \\ (0,1), & \text{if parity} = 2 \text{ for the } i^{th} \text{ record} \\ (0,0), & \text{otherwise} \end{cases}$$

and

$$w_{i,1103} = \begin{cases} 1, & \text{if the } i^{th} \text{ calf is female} \\ 0, & \text{if the } i^{th} \text{ calf is male} \end{cases} .$$

The system of equations (3.51) for computing $\underline{\tau}^{(k)}$ and $\underline{s}^{(k)}$ have order 1188. Rather than constructing this entire system, it is advantageous to 'absorb' the HYS equations, i.e. the first 1100 equations, into the other 88 equations to obtain a system of equations involving only the parity, sex, and sire effects. After the solution is computed for this reduced system of equations, the solution for each

HYS effect is easily obtained, since each of the HYS equations involve only 1 HYS effect.

We computed estimates for the parameters of the alternative mixed effects threshold model as described in Section 4. The estimates for the boundaries, the parity and sex effects, and the variance ratio are given in Table 4 [τ_{1101} = (Parity = 1) - (Parity $\geq$ 3), τ_{1102} = (Parity = 2) - (Parity $\geq$ 3), τ_{1103} = (Female - Male)].

Table 4. Estimates for Parameters of the Alternative Threshold Model

Parameter	Estimate 1st A	Estimate 2nd A
η_1	.5453	.5243
η_2	.9254	.9046
η_3	1.6436	1.6231
η_4	2.2878	2.2675
τ_{1101}	.9939	.9958
τ_{1102}	.1164	.1165
τ_{1103}	-.4097	-.4106
ρ_1	.0373	.0382

Heritability (h^2), a quantity of theoretical importance to animal breeders, is defined to be the ratio of additive genetic variance to total (phenotypic) variance, and may be estimated by $4\hat{\rho}_1/(1+\hat{\rho}_1)$, i.e., by 4 times the estimated correlation between calvings from the same sire (Falconer, 1960, pg. 169). The estimate of h^2 using the 2nd A is

.147. (Using the 1st A, h^2 is estimated to be .144.) Previous estimates of h^2 for calving difficulty have varied considerably with parity and with breed. Pollak's (1975) estimates for the heritability of calving difficulty among Holsteins were .17, .08, and .05 for first, second, and third or greater parities, respectively.

We now describe the model for calving difficulty assumed by Berger and Freeman (B & F) (1978) and then compare the analysis of the ABS data based on their model with the threshold model analysis.

The variable Z_i (as previously defined) with possible values 1,2, ...,5 is assumed to satisfy a mixed linear model which includes fixed effects for HYS, parity, and sex and a random effect for sires. As with the threshold model, the covariance for the sire effects is taken as $\sigma_s^2\underline{A}_1$, where $\underline{A}_1$ is the relationship matrix and σ_s^2 is an unknown parameter, but the covariance for the errors is assumed to be $\sigma_e^2\underline{R}$, where $\underline{R}$ is diagonal with $(i,i)^{th}$ element r_i given by

$$r_i = \begin{cases} 1.44, & \text{if parity (for dam of } i^{th} \text{ calf)} = 1 \\ .78, & \text{if parity (for dam of } i^{th} \text{ calf)} = 2 \\ .65, & \text{if parity (for dam of } i^{th} \text{ calf)} \geq 3 \end{cases} .$$

(The relatively infrequent occurrence of calving difficulty for second or higher parities implies a lower variance for these parities.) The elements of $\underline{R}$ are based on estimates reported by Pollak (1975).

Estimates for σ_s^2 and σ_e^2 were computed iteratively using equations (6.3) and (6.4) of Harville (1977). These equations are for computing REML estimates of variance components (assuming the random effects and errors are normally distributed). We obtained

$$\sigma_s^2 = .0077 \tag{3.61}$$

and

$$\sigma_e^2 = .8605 .$$

[Berger and Freeman (1978) obtained estimates for σ_s^2 and σ_e^2 equal to .026 and .860, respectively, for the data they analyzed.]

Since this model assumes that the sire effect variance is constant across parities, it implies that $\sigma_s^2/(r_i\sigma_e^2)$ and, hence, h^2 are increasing functions of parity. This seems undesirable, since Pollak's (1975) estimates for this ratio decrease with parity.

We now compare the estimates of the sire effects based on the threshold analysis with the estimates obtained using the B & F model (refer to Table 5). Predictions (PRED) of the sire effects are listed for: 1) Threshold model, 1st A, 2) Threshold model, 2nd A, 3) Berger and Freeman model, $\sigma_s^2/\sigma_e^2 = .009$ [refer to (3.61)], and 4) Berger and Freeman model, $\sigma_s^2/\sigma_e^2 = .030$ (their estimate for this ratio). The 85 sires are identified by their registration numbers (SIRE ID). R denotes the rank for each of the 4 procedures. The sires are listed in the order they were ranked by the threshold analysis, 1st A. The estimated standard error (ST ERR) of prediction is given for each PRED.

N1 denotes the number of records for each sire, N2 denotes the number of records for each sire which have comparisons with other sires within a HYS, and N3 denotes the number of records for each sire which have comparisons with other sires within a HYS, after the 2975 records (contained in HYS's for which all the response were in Category 1 or all in Category 5) were deleted. N2 represents the number of records for each sire useful for making comparisons between sires under the B & F model. N3 represents the corresponding number for the threshold model. Although N2 $\geq$ N3, the records represented in N2, but not in N3, are used by the B & F procedure (which scores the responses) because it includes the assumption that all of the responses in, e.g., the first category are identical. As mentioned in Section 1, this is an undesirable assumption.

Using the threshold model, 1st A, we estimate, for each sire, the probability of a difficult birth for heifers. We computed the probability of a difficult birth for each sex, and then, took the simple average of these probabilities. The HYS effect was chosen so that the probability of a response in Category 4 or 5 would be 13-14%. This is to represent a 'typical' HYS. The value used was -.245 (this quantity representing the difference between the first HYS and the typical HYS). We provide both the probability of a response in Category 4 or 5 (PR>3) and the probability of a response in Category 5 (PR>4).

We illustrate the calculations for the sire ranked first. Given that the calf is a male, PR>3 is estimated by

$$1 - \Phi\{[1.6436 - (-.245 + .9939 - .3369)]/[1 + (.0755)^2]^{\frac{1}{2}}\}$$

which equals .100. Given that the calf is a female, PR>3 is .051. The average PR>3, as reported in Table 5, for the first sire is .076.

In conjunction with the B & F approach, the frequency of births from heifers scored 4 or 5 (FR>3) for each sire, may be estimated by

$$(FR>3) = .1387 + .56(PRED). \qquad (3.62)$$

[Equation (3.62) was obtained from a previous analysis by a simple linear regression of the observed frequency of births rated 4 or 5 on the sire estimates.] The estimated FR>3, calculated using (3.62), is given for the B & F analyses. (The estimated FR>3 is not the same as PR>3, since FR>3 depends only on PRED and not on the ST ERR.) We note that linear models for frequencies often fit poorly for values close to 0 or close to 1.

6. DISCUSSION

We make several observations about the computations for the estimation of the parameters of the threshold model. For each k(i.e., $\underline{\rho}^{(k)}$), Step 1' involves the iterative computation of $\hat{\underline{\tau}}^{(k)}$, $\hat{\underline{s}}^{(k)}$, and $\hat{\underline{\eta}}^{(k)}$ using systems (3.51) and (3.57). Rather than recomputing the coefficient matrices of these two systems at each new value of $\underline{\eta}$, $\underline{\tau}$, and $\underline{s}$, we only recomputed these matrices each time the value of $\underline{\rho}$ was recomputed. This approach seemed to work effectively.

For the ABS data, we started with $\underline{\tau}^{(0)} = \underline{0}$, $\underline{s}^{(0)} = 0$ and with an initial value for $\underline{\eta}$ as described in Mee and Harville (1981b) for the fixed effects threshold model. Computation of $\hat{\underline{\tau}}^{(1)}$, $\hat{\underline{s}}^{(1)}$, and $\hat{\underline{\eta}}^{(1)}$ required 7-8 iterations (and approximately 18 seconds CPU time for the ITEL AS/6). Computing $\rho_1^{(2)}$ and recomputing the coefficient matrices of (3.51) and (3.57) required approximately 9 seconds CPU time. Computation of $\hat{\underline{\tau}}^{(k)}$, $\hat{\underline{s}}^{(k)}$, and $\hat{\underline{\eta}}^{(k)}$ on subsequent iterations, i.e., Step 1' for k = 2,3,... required 5 seconds or less, since the change from $\hat{\underline{\tau}}^{(k-1)}$ to $\hat{\underline{\tau}}^{(k)}$, etc., was generally small. The computing time was essentially the same for the 1st A as for the 2^{nd} A.

We started the computations at several trial values for $\rho_1^{(1)}$ to determine the approximate value of $\hat{\rho}_1$. Then we began with $\rho_1^{(1)} = .037$ and, iterating until $|\rho_1^{(k+1)} - \rho_1^{(k)}| < .0005$, (for the 1^{st} A) obtained $\rho_1^{(2)} = .0364$, $\rho_1^{(3)} = .0369$, and lastly $\rho_1^{(4)} = .0373$. [The decrease from $\rho_1^{(1)}$ to $\rho_1^{(2)}$ was surprising since, in the continuous case, the EM algorithm has often shown monotonic convergence for variance components. The initial drop in ρ_1 using our approximation to the EM algorithm may possibly be attributed to the value of $\hat{\underline{\Sigma}}_1^{(1)}$ used in (3.60).] For the 2nd A, we set $\rho_1^{(1)} = .037$ and, iterating for 5 iterations, obtained .0364, .0370, .0375, .0379, and .0382 for $\rho_1^{(k)}$, k = 2,...,6.

The 1st A and 2nd A gave identical rankings for sires with the Quass and Van Vleck data and virtually identical rankings with the ABS data. However, in both analyses, the estimate for ρ_1 was slightly higher when the 2nd A was being used (i.e., .1723 vs. .1702 and .0382 vs. .0373).

Using the B & F model, the estimates [refer to (3.61)] for the ratio of the sire variance to error variance $(r_1\sigma_e^2)$ were .0062, .0114, and .0137 for parities 1, 2, and 3 or greater. These values are much below the estimate of this ratio based on the threshold model.

Table 5. Comparison of Predictions

					THRESHOLD MODEL						
					FIRST APPROX				SECOND APPROX		
R	SIRE ID	N1	N2	N3	PRED	ST ERR	PR>3	PR>4	R	PRED	ST ERR
1	1563679	866	801	578	-0.3369	0.0755	0.076	0.020	1	-0.3371	0.0758
2	1593395	172	156	121	-0.2149	0.1255	0.096	0.027	2	-0.2189	0.1264
3	1564328	1621	1505	1206	-0.2116	0.0625	0.095	0.026	3	-0.2106	0.0629
4	1575285	424	351	241	-0.1790	0.0960	0.101	0.028	4	-0.1799	0.0964
5	1647459	311	306	253	-0.1674	0.1015	0.103	0.029	5	-0.1688	0.1021
6	1693308	42	37	29	-0.1508	0.1643	0.108	0.031	6	-0.1548	0.1659
7	1613580	147	143	117	-0.1470	0.1132	0.107	0.031	7	-0.1489	0.1139
8	1564147	701	540	415	-0.1055	0.0874	0.114	0.033	10	-0.1048	0.0878
9	1648465	36	31	26	-0.1051	0.1599	0.116	0.034	8	-0.1077	0.1615
10	1684833	35	33	27	-0.1026	0.1677	0.117	0.035	9	-0.1055	0.1694
11	1672322	29	26	20	-0.0977	0.1689	0.118	0.035	11	-0.1005	0.1706
12	1608369	48	37	29	-0.0902	0.1601	0.119	0.036	12	-0.0929	0.1616
13	1694079	30	27	16	-0.0707	0.1663	0.123	0.037	13	-0.0735	0.1681
14	1664494	45	34	29	-0.0673	0.1585	0.123	0.037	14	-0.0687	0.1600
15	1617071	504	491	423	-0.0648	0.0826	0.122	0.036	15	-0.0641	0.0830
16	1648032	8	8	7	-0.0546	0.1839	0.127	0.039	16	-0.0568	0.1861
17	1635790	48	48	42	-0.0542	0.1526	0.126	0.038	17	-0.0563	0.1539
18	1663500	13	10	10	-0.0524	0.1781	0.127	0.039	18	-0.0546	0.1801
19	1635912	125	123	115	-0.0508	0.1188	0.125	0.038	20	-0.0511	0.1194
20	1633380	55	55	45	-0.0500	0.1479	0.126	0.039	19	-0.0524	0.1493
21	1672519	7	7	4	-0.0356	0.1880	0.131	0.041	21	-0.0371	0.1903
22	1658837	15	10	5	-0.0255	0.1735	0.132	0.041	22	-0.0268	0.1755
23	1629391	249	231	193	-0.0241	0.1047	0.131	0.040	23	-0.0244	0.1053
24	1669038	3	3	0	-0.0202	0.1860	0.134	0.042	24	-0.0213	0.1884
25	1678178	26	19	14	-0.0193	0.1769	0.134	0.042	25	-0.0197	0.1789
26	1658167	8	8	7	-0.0185	0.1840	0.134	0.042	26	-0.0194	0.1862
27	1599189	67	64	52	-0.0171	0.1448	0.133	0.041	27	-0.0179	0.1460
28	1602410	329	321	257	-0.0150	0.0947	0.132	0.041	28	-0.0148	0.0951
29	1672121	19	11	5	-0.0127	0.1816	0.135	0.042	30	-0.0137	0.1837
30	1642993	63	62	42	-0.0122	0.1424	0.134	0.042	29	-0.0138	0.1437
31	1660592	14	11	10	-0.0115	0.1749	0.135	0.042	31	-0.0124	0.1768
32	1695206	35	27	21	-0.0111	0.1635	0.135	0.042	32	-0.0119	0.1651
33	1673127	14	9	8	-0.0079	0.1720	0.136	0.043	33	-0.0086	0.1740
34	1684205	27	21	21	-0.0065	0.1605	0.136	0.042	34	-0.0075	0.1621
35	1673806	2	2	0	-0.0038	0.1793	0.137	0.043	35	-0.0043	0.1816
36	1693030	3	3	3	-0.0027	0.1757	0.137	0.043	36	-0.0036	0.1778
37	1658701	7	3	3	-0.0003	0.1839	0.138	0.044	37	-0.0012	0.1861
38	1678710	7	7	6	0.0016	0.1780	0.138	0.044	38	0.0015	0.1801
39	1672151	1	1	0	0.0025	0.1794	0.138	0.044	40	0.0022	0.1816
40	1698738	31	26	21	0.0031	0.1631	0.138	0.043	39	0.0016	0.1648
41	1685359	24	21	18	0.0033	0.1668	0.138	0.043	41	0.0023	0.1685
42	1689994	22	22	19	0.0054	0.1689	0.139	0.044	42	0.0050	0.1707

DEL			BERGER AND FREEMAN MODEL							
SECOND APPROX			$\sigma_s^2/\sigma_e^2 = .009$				$\sigma_s^2/\sigma_e^2 = .030$			
R	PRED	ST ERR	R	PRED	ST ERR	FR>3	R	PRED	ST ERR	FR>3
1	-0.3371	0.0758	1	-0.1665	0.0376	0.045	1	-0.1900	0.0463	0.032
2	-0.2189	0.1264	4	-0.0757	0.0587	0.096	3	-0.1260	0.0747	0.068
3	-0.2106	0.0629	2	-0.0993	0.0313	0.083	4	-0.1087	0.0404	0.078
4	-0.1799	0.0964	3	-0.0771	0.0479	0.096	5	-0.1005	0.0582	0.082
5	-0.1688	0.1021	5	-0.0750	0.0478	0.097	7	-0.0973	0.0584	0.084
6	-0.1548	0.1659	6	-0.0598	0.0776	0.105	2	-0.1295	0.1162	0.066
7	-0.1489	0.1139	7	-0.0584	0.0568	0.106	6	-0.0977	0.0725	0.084
10	-0.1048	0.0878	8	-0.0521	0.0426	0.110	14	-0.0624	0.0516	0.104
8	-0.1077	0.1615	10	-0.0414	0.0752	0.116	8	-0.0867	0.1133	0.090
9	-0.1055	0.1694	15	-0.0320	0.0786	0.121	10	-0.0761	0.1203	0.096
11	-0.1005	0.1706	11	-0.0378	0.0789	0.118	9	-0.0797	0.1207	0.094
12	-0.0929	0.1616	12	-0.0362	0.0765	0.118	11	-0.0752	0.1125	0.097
13	-0.0735	0.1681	14	-0.0339	0.0747	0.120	12	-0.0707	0.1139	0.099
14	-0.0687	0.1600	33	-0.0073	0.0753	0.135	43	0.0032	0.1103	0.140
15	-0.0641	0.0830	9	-0.0456	0.0411	0.113	16	-0.0577	0.0505	0.106
16	-0.0568	0.1861	22	-0.0185	0.0841	0.128	18	-0.0505	0.1409	0.110
17	-0.0563	0.1539	40	0.0003	0.0726	0.139	39	0.0021	0.1020	0.140
18	-0.0546	0.1801	35	-0.0042	0.0829	0.136	32	-0.0208	0.1376	0.127
20	-0.0511	0.1194	31	-0.0086	0.0650	0.134	34	-0.0146	0.0847	0.131
19	-0.0524	0.1493	13	-0.0342	0.0688	0.120	13	-0.0625	0.0964	0.104
21	-0.0371	0.1903	30	-0.0098	0.0852	0.133	22	-0.0362	0.1472	0.118
22	-0.0268	0.1755	16	-0.0270	0.0799	0.124	15	-0.0613	0.1345	0.104
23	-0.0244	0.1053	32	-0.0080	0.0544	0.134	28	-0.0247	0.0672	0.125
24	-0.0213	0.1884	47	0.0066	0.0843	0.142	51	0.0158	0.1458	0.148
25	-0.0197	0.1789	37	-0.0025	0.0810	0.137	52	0.0177	0.1291	0.149
26	-0.0194	0.1862	20	-0.0192	0.0848	0.128	17	-0.0567	0.1451	0.107
27	-0.0179	0.1460	43	0.0024	0.0717	0.140	41	0.0023	0.0995	0.140
28	-0.0148	0.0951	34	-0.0056	0.0468	0.136	33	-0.0161	0.0575	0.130
30	-0.0137	0.1837	50	0.0078	0.0832	0.143	49	0.0140	0.1399	0.147
29	-0.0138	0.1437	29	-0.0098	0.0671	0.133	20	-0.0425	0.0945	0.115
31	-0.0124	0.1768	46	0.0062	0.0833	0.142	46	0.0105	0.1390	0.145
32	-0.0119	0.1651	26	-0.0108	0.0789	0.133	25	-0.0297	0.1214	0.122
33	-0.0086	0.1740	17	-0.0227	0.0800	0.126	19	-0.0439	0.1349	0.114
34	-0.0075	0.1621	23	-0.0175	0.0764	0.129	21	-0.0407	0.1205	0.116
35	-0.0043	0.1816	24	-0.0152	0.0818	0.130	29	-0.0227	0.1446	0.126
36	-0.0036	0.1778	27	-0.0105	0.0807	0.133	31	-0.0208	0.1398	0.127
37	-0.0012	0.1861	56	0.0162	0.0848	0.148	60	0.0336	0.1476	0.158
38	0.0015	0.1801	38	-0.0017	0.0818	0.138	45	0.0086	0.1375	0.144
40	0.0022	0.1816	25	-0.0138	0.0819	0.131	30	-0.0214	0.1450	0.127
39	0.0016	0.1648	53	0.0110	0.0761	0.145	37	-0.0075	0.1171	0.134
41	0.0023	0.1685	63	0.0246	0.0790	0.152	67	0.0476	0.1228	0.165
42	0.0050	0.1707	44	0.0035	0.0803	0.141	38	0.0006	0.1266	0.139

Table 5 (Continued)

					THRESHOLD MODEL						
					FIRST APPROX				SECOND APPROX		
R	SIRE ID	N1	N2	N3	PRED	ST ERR	PR>3	PR>4	R	PRED	ST ERR
43	1695218	21	21	10	0.0056	0.1771	0.139	0.044	43	0.0062	0.179
44	1636562	43	43	34	0.0070	0.1536	0.138	0.043	44	0.0064	0.155
45	1589857	552	508	426	0.0081	0.0844	0.137	0.042	45	0.0084	0.084
46	1636144	48	48	45	0.0103	0.1490	0.139	0.044	46	0.0097	0.150
47	1670654	4	4	4	0.0124	0.1774	0.141	0.045	47	0.0115	0.179
48	1601125	62	62	52	0.0160	0.1491	0.140	0.044	48	0.0153	0.150
49	1639748	2	2	2	0.0187	0.1870	0.142	0.045	49	0.0189	0.189
50	1689995	10	10	8	0.0190	0.1785	0.142	0.045	50	0.0190	0.180
51	1653334	1	1	1	0.0206	0.1825	0.142	0.045	51	0.0201	0.184
52	1690469	44	41	28	0.0209	0.1523	0.141	0.045	53	0.0216	0.153
53	1686245	32	32	20	0.0211	0.1573	0.142	0.045	52	0.0210	0.158
54	1697162	21	20	15	0.0230	0.1755	0.143	0.045	54	0.0226	0.177
55	1680121	8	8	6	0.0339	0.1711	0.145	0.046	56	0.0340	0.173
56	1692150	22	22	14	0.0347	0.1705	0.145	0.046	55	0.0339	0.172
57	1692619	39	39	24	0.0394	0.1629	0.146	0.047	57	0.0387	0.164
58	1616359	136	129	118	0.0409	0.1190	0.145	0.046	58	0.0408	0.119
59	1650414	30	30	25	0.0472	0.1553	0.147	0.047	59	0.0470	0.156
60	1671167	14	10	9	0.0474	0.1719	0.148	0.048	60	0.0479	0.173
61	1671336	24	10	7	0.0486	0.1725	0.148	0.048	61	0.0485	0.174
62	1669592	8	6	6	0.0539	0.1739	0.150	0.048	62	0.0544	0.176
63	1634703	20	20	20	0.0556	0.1627	0.149	0.048	63	0.0559	0.164
64	1589706	701	645	506	0.0584	0.0791	0.148	0.047	64	0.0599	0.079
65	1686926	4	4	1	0.0626	0.1889	0.152	0.050	65	0.0652	0.191
66	1615951	363	361	312	0.0704	0.0903	0.151	0.048	66	0.0718	0.090
67	1680975	49	38	27	0.0767	0.1596	0.154	0.050	67	0.0770	0.161
68	1693040	5	5	1	0.0794	0.1699	0.155	0.051	68	0.0802	0.172
69	1659046	5	5	2	0.0876	0.1754	0.157	0.052	69	0.0891	0.177
70	1638034	67	67	63	0.0895	0.1365	0.156	0.051	70	0.0905	0.137
71	1631223	354	342	296	0.1022	0.0898	0.158	0.051	71	0.1039	0.090
72	1647417	9	9	8	0.1064	0.1778	0.162	0.054	72	0.1087	0.179
73	1698093	34	34	28	0.1116	0.1639	0.162	0.054	73	0.1134	0.165
74	1652465	99	90	81	0.1246	0.1248	0.164	0.054	74	0.1258	0.125
75	1697421	35	24	18	0.1266	0.1640	0.166	0.056	75	0.1284	0.165
76	1518703	579	493	384	0.1374	0.0864	0.166	0.055	76	0.1395	0.086
77	1450228	330	283	214	0.1408	0.0916	0.167	0.056	77	0.1431	0.092
78	1682256	25	20	14	0.1526	0.1713	0.173	0.059	78	0.1550	0.173
79	1594937	359	342	292	0.1538	0.0903	0.171	0.057	79	0.1558	0.090
80	1593263	1380	1305	1095	0.1641	0.0619	0.173	0.058	80	0.1663	0.062
81	1632079	61	59	49	0.1642	0.1399	0.174	0.059	81	0.1666	0.141
82	1608372	475	459	408	0.2189	0.0835	0.187	0.065	82	0.2214	0.083
83	1629385	715	678	568	0.2328	0.0731	0.191	0.066	83	0.2352	0.073
84	1603894	271	261	226	0.2590	0.1004	0.198	0.070	84	0.2619	0.100
85	1633540	58	58	51	0.3395	0.1330	0.221	0.082	85	0.3446	0.133

)EL			BERGER AND FREEMAN MODEL							
SECOND APPROX			$\sigma_s^2/\sigma_e^2=.009$				$\sigma_s^2/\sigma_e^2=.030$			
R	PRED	ST ERR	R	PRED	ST ERR	FR>3	R	PRED	ST ERR	FR>3
43	0.0062	0.1792	28	-0.0099	0.0789	0.133	44	0.0049	0.1253	0.141
44	0.0064	0.1550	19	-0.0195	0.0724	0.128	26	-0.0285	0.1063	0.123
45	0.0084	0.0847	18	-0.0223	0.0486	0.126	23	-0.0330	0.0589	0.120
46	0.0097	0.1503	49	0.0076	0.0713	0.143	55	0.0213	0.1004	0.151
47	0.0115	0.1795	51	0.0086	0.0824	0.144	42	0.0027	0.1410	0.140
48	0.0153	0.1504	58	0.0221	0.0692	0.151	65	0.0464	0.0951	0.165
49	0.0189	0.1893	52	0.0101	0.0870	0.144	58	0.0305	0.1562	0.156
50	0.0190	0.1806	48	0.0075	0.0829	0.143	50	0.0146	0.1376	0.147
51	0.0201	0.1647	54	0.0122	0.0841	0.146	48	0.0126	0.1480	0.146
53	0.0216	0.1537	21	-0.0191	0.0722	0.128	36	-0.0090	0.1061	0.134
52	0.0210	0.1589	39	-0.0007	0.0720	0.138	27	-0.0282	0.1095	0.123
54	0.0226	0.1775	42	0.0014	0.0800	0.139	35	-0.0106	0.1278	0.133
56	0.0340	0.1731	36	-0.0031	0.0795	0.137	40	0.0021	0.1339	0.140
55	0.0339	0.1724	62	0.0238	0.0780	0.152	62	0.0341	0.1220	0.158
57	0.0387	0.1646	60	0.0227	0.0743	0.151	57	0.0303	0.1092	0.156
58	0.0408	0.1197	66	0.0274	0.0595	0.154	56	0.0264	0.0758	0.153
59	0.0470	0.1567	68	0.0290	0.0753	0.155	68	0.0522	0.1118	0.168
60	0.0479	0.1738	57	0.0186	0.0798	0.149	63	0.0393	0.1329	0.161
61	0.0485	0.1745	45	0.0043	0.0800	0.141	24	-0.0307	0.1333	0.122
62	0.0544	0.1760	41	0.0011	0.0801	0.139	47	0.0122	0.1369	0.146
63	0.0559	0.1643	64	0.0250	0.0788	0.153	66	0.0471	0.1239	0.165
64	0.0599	0.0794	59	0.0223	0.0394	0.151	54	0.0199	0.0482	0.150
65	0.0652	0.1913	65	0.0267	0.0855	0.154	78	0.1219	0.1497	0.207
66	0.0718	0.0907	72	0.0392	0.0446	0.161	64	0.0452	0.0543	0.164
67	0.0770	0.1610	74	0.0412	0.0761	0.162	74	0.0783	0.1140	0.183
68	0.0802	0.1720	67	0.0280	0.0774	0.154	61	0.0340	0.1334	0.158
69	0.0891	0.1775	61	0.0232	0.0805	0.152	75	0.0817	0.1387	0.184
70	0.0905	0.1376	70	0.0302	0.0667	0.156	70	0.0662	0.0909	0.176
71	0.1039	0.0902	55	0.0149	0.0450	0.147	53	0.0185	0.0549	0.149
72	0.1087	0.1798	75	0.0429	0.0842	0.163	79	0.1237	0.1415	0.208
73	0.1134	0.1655	73	0.0405	0.0776	0.161	76	0.0863	0.1159	0.187
74	0.1258	0.1256	71	0.0353	0.0634	0.158	69	0.0581	0.0838	0.171
75	0.1284	0.1657	77	0.0602	0.0775	0.172	81	0.1295	0.1188	0.211
76	0.1395	0.0868	69	0.0293	0.0450	0.155	59	0.0331	0.0544	0.157
77	0.1431	0.0921	76	0.0477	0.0456	0.165	72	0.0710	0.0593	0.178
78	0.1550	0.1732	81	0.0741	0.0793	0.180	84	0.1725	0.1255	0.235
79	0.1558	0.0907	78	0.0647	0.0457	0.175	73	0.0758	0.0560	0.181
80	0.1663	0.0622	79	0.0684	0.0312	0.177	71	0.0698	0.0404	0.178
81	0.1666	0.1410	80	0.0692	0.0706	0.177	83	0.1393	0.0986	0.217
82	0.2214	0.0839	82	0.0856	0.0431	0.187	77	0.0975	0.0525	0.193
83	0.2352	0.0734	84	0.1184	0.0379	0.205	82	0.1310	0.0468	0.212
84	0.2619	0.1009	83	0.1024	0.0493	0.196	80	0.1291	0.0607	0.211
85	0.3446	0.1339	85	0.1363	0.0676	0.215	85	0.2544	0.0958	0.281

7. REFERENCES

Aitkin, M. 1981. Private correspondence. University of Lancaster.

Berger, P.J. and Freeman, A.E. 1978. Prediction of sire merit for calving difficulty. Journal of Dairy Science 61, 1146-1150.

Burridge, J. 1981. A note on maximum likelihood estimation for regression models using grouped data. Journal of the Royal Statistical Society Ser. B 39, 1-38.

Dempster, A.P., Laird, N.M., and Rubin, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Ser B 39, 1-38.

Dempster, A.P., Rubin, D. B., and Tsutakawa, R.K. 1981. Estimation in covariance components models. Journal of the American Statistical Association 76, 341-353.

Falconer, D.S. 1960. Introduction to Quantitative Genetics. New York: Ronald Press Co.

Gianola, D. 1980. Genetic evaluation of animals for traits with categorical responses. Journal of Animal Science 41, 1272-1276.

Harville, D.A. 1976. Extension of the Gauss-Markov theorem to include the estimation of random effects. Annals of Statistics 4, 384-395.

Harville, D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association 72, 320-340.

Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31, 423-447.

Laird, N.M. 1975. Log-linear models with random parameters: an empirical Bayes approach. Ph.D. dissertation. Harvard University, Cambridge.

Landis, J.R. and Koch, G.G. 1977. A one-way components of variance model for categorical data. Biometrics 33, 671-679.

Mee, R.W. and Harville, D.A. 1981a. Analysis of ordered categorical responses, assuming an underlying continuous variable. Submitted to the Journal of the American Statistical Association.

Mee. R.W. and Harville, D.A. 1981b. Analysis of ordinal data via the threshold model. Submitted to Biometrics.

Owen, D.B. 1956. Tables for computing bivariate normal probabilities. Annals of Mathematical Statistics 27, 1075-1090.

Pollak, E.J. 1975. Dystocia in Holsteins. Ph.D. dissertation. Iowa State University, Ames.

Quass, R.L. and Van Vleck, L.D. 1980. Categorical trait sire evaluation by best linear unbiased prediction of future progeny category frequency. Biometrics 36, 117-122.

Searle, S.R. 1971. Linear Models. New York: Wiley.

Tong, A.K.W., Wilton, J.W. and Schaeffer, L.R. 1977. Application of a scoring procedure and transformations to dairy type classification and beef ease of calving categorical data. Canadian Journal of Animal Science 57, 1-5.

8. APPENDIX

Suppose the linear model (3.16) for t_i is simply

$$t_i = s + e_i,$$

where $Var(s)=\rho_1$. We consider three cases:

Case	η_1	η_2	Number of observations (by category) 1	2	3
1	0	1	0	1	0
2	-3	-2	0	0	1
3	2	3	10	6	2

For each case, we compute the mean and variance of s, conditional on the categorical observations, e.g., for Case 1, we compute $E[s|\ 0< t_1 <1]$, etc. In cases with a single observation, the conditional mean and variance may be computed analytically. For Case 3, however, the 'exact' values had to be computed numerically. We compare the exact values with those obtained using the approximations (3.45) and (3.55) (we refer to these as the 1^{st} A and the 2^{nd} A, respectively) for 5 values of ρ_1: .01, .1, .5, 4, and 100. The $Var[s|data]$ is approximated using (3.52) evaluated at the 2^{nd} A value for $E[s|data]$. The results appear in Table A1.

The approximations for the conditional mean of s are close to the exact values for Cases 1 and 3. For Case 2, the 1^{st} A is poor, except when ρ_1 is very small. The approximation for the variance is best for Case 3 (where there are several responses) and poorest for Case 2. Unless all the responses (corresponding to a certain random effect) are in a single extreme category, we expect both approximations to be adequate.

Table A1. Comparisons of Approximations with Exact Values

Case	ρ_1	E[s|data]			Var[s|data]	
		Exact	2^{nd} A	1^{st} A	Exact	2^{nd} A
1	.01	.0046	.0046	.0046	.0099	.0099
	.1	.0421	.0421	.0421	.0916	.0929
	.5	.1576	.1576	.1575	.3423	.3594
	4.0	.3934	.3933	.3930	.8530	.9511
	100.0	.4946	.4944	.4944	1.0724	1.2293
2	.01	.0006	.0006	.0006	.0100	.0100
	.1	.0064	.0063	.0054	.0988	.0987
	.5	.0452	.0448	.0255	.4678	.4710
	4.0	.5874	.5994	.1514	2.7153	3.4458
	100.0	6.7230	2.2612	.3768	41.4880	97.9315
3	.01	.1830	.1828	.1829	.0092	.0097
	.1	1.0049	1.0041	1.0077	.0475	.0548
	.5	1.5847	1.5833	1.5894	.0688	.0721
	4.0	1.7969	1.7963	1.8025	.0750	.0760
	100.0	1.8298	1.8292	1.8355	.0760	.0766

SUMMARY AND DISCUSSION

This dissertation contributes to the analysis of ordered categorical responses based on the threshold appoach, particularly in the extension to the case where random effects are included in the assumed linear model for the underlying continuous response variable. Although approximations are required in the estimation procedures, the approximations will generally be accurate. Confidence bounds for probabilities based on the mixed effects threshold model could be constructed by procedures similar to those described in Part I, Section 4, for the fixed effects threshold model.

It would be useful to further investigate the appropriateness of the model we have assumed for the analysis of calving difficulty. One question is whether the assumption of a constant (with respect to parity) sire variance to error variance ratio is reasonable. A second question is whether the estimate for the sire variance to error variance ratio and the ranking of (well-estimated) sires based on this model will be relatively consistent from one analysis to another.

Other extensions of the threshold approach could be pursued. Consider the situation of a bivariate response. When one observed response is continuous and the other is ordered categorical, the estimation procedures could be developed based on the EM algorithm (and certain approximations) similar to the development of Part III. (The case where both observed responses are ordered categorical would be much more difficult.)

LITERATURE CITED

Ashford, J.R. 1959. An approach to the analysis of data for semi-quantal responses in biological assay. Biometrics 15, 573-581.

Berger, P.J. and Freeman, A.E. 1978. Prediction of sire merit for calving difficulty. Journal of Dairy Science 61, 1146-1150.

Bock, R. D. 1975. Multivariate Statistical Methods in Behavioral Research. New York: McGraw Hill, Inc.

Falconer, D.S. 1960. Introduction to Quantitative Genetics. New York: Ronald Press Co.

Gianola, D. 1979. Heritability of polychotomous characters. Genetics 93, 1051-1055.

Gurland, J., Lee, I., and Dahm, P.A. 1960. Polychotomous quantal response in biological assay. Biometrics 16, 382-398.

Laird, N.M. 1975. Log-linear models with random parameters: an emperical Bayes approach. Ph.D. dissertation. Harvard Univerity, Cambridge.

Landis, J.R. and Koch, G.G. 1977. A one-way components of variance model for categorical data. Biometrics 33, 671-679.

McCullagh, P. 1980. Regression models for ordinal data. Journal of the Royal Statistical Society Ser. B 42, 109-127.

McKelvey, R.D. and Zavoina, W. 1975. A statistical model for the analysis of ordinal level dependent variables. Journal of Mathematical Sociology 4, 103-120.

Quass, R.L. and Van Vleck, L.D. 1980. Categorical trait sire evaluation by best linear unbiased prediction of future progeny category frequency. Biometrics 36, 117-122.

Snell, E.J. 1964. A scaling procedure for ordered categorical data. Biometrics 20, 592-607.

Thompson, R. 1979. Sire evaluation. Biometrics 35, 339-353.

Tong, A.K.W., Wilton, J.W. and Schaeffer, L.R. 1976. Evaluation of ease of calving for Charolais sires. Canadian Journal of Animal Science 56, 17-26.

Tong, A.K.W., Wilton, J.W. and Schaeffer, L.R. 1977. Application of a scoring procedure and transformations to dairy type classification and beef ease of calving categorical data. Canadian Journal of Animal Science 57, 1-5.

ACKNOWLEDGEMENTS

"O Lord, thou wilt ordain peace for us,

thou hast wrought for us all our works."(Isaiah 26: 12)

I acknowledge with thanksgiving that God has fulfilled the promise I believed four years ago regarding my study. I thank him for granting success and for the strength and stability of his presence in me.

I greatly appreciate the direction David Harville gave to me in this work. He saw so much from the outset. His demands for thoroughness from me were fair and needed, and I have benefitted. I particularly thank him for his patience and his instruction as I have attempted to write.

Other faculty members also deserve recognition. I wish to thank Eugene Freeman, Wayne Fuller, Malay Ghosh, and Kenneth Koehler for serving as my committee members. Special thanks is due to Kenneth Koehler, for I sought and received help from him on many occasions.

I thank Peter Hoffman and Robert Brennan of the Department of Animal Science for the nutrition data used in Part II. Jeffrey Berger provided me with the American Breeder Service data and assisted me with the computer program for the analysis of those data. I thank him for his availability to me the many times I went to him for help.

My parents have given me continual encouragement. I am rich through their loving support and prayers.

This work was supported by the Air Force Office of Scientific Research under Grant AROSR 76-3037.